ALL-ATOM MODELING FOR SOLUTION-PROCESSED SOLAR CELL MATERIALS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

James Minuse Stevenson

August 2016

© 2016 James Minuse Stevenson

ALL RIGHTS RESERVED

ALL-ATOM MODELING FOR SOLUTION-PROCESSED SOLAR CELL MATERIALS

James Minuse Stevenson, Ph.D.

Cornell University 2016

The field of photovoltaics is in the midst of a revolution provoked by hybrid metal and organic solar cell materials, which combine some of the best properties of cheap but inefficient organic solar cells and relatively expensive but efficient metallic cells. The key to these new hybrid materials is that their synthesis takes place in organic liquid solutions, and yet produces effective ordered solid solar cells. However, the synthesis methods which produce these results have been discovered almost by accident, and it is not clear how to improve them. In this work I propose a broad strategy for understanding these processes based on all-atom simulation via Density Functional Theory and classical force fields. With these methods I elucidate some of the fundamental mechanisms behind the success of metalorganic solar cells, and show a way forward to larger-scale production for these new technologies.

BIOGRAPHICAL SKETCH

James Stevenson is a scientist, engineer, and activist from New York City. As part of his PhD work at Cornell with Prof. Paulette Clancy, James has researched a wide variety of topics, including solution-processed solar cells, methane clathrate hydrates, and life on Saturn's moons. Prior to his work at Cornell, James was an undergraduate at The Cooper Union for the Advancement of Science and Art under the advisement of Prof. Benjamin Davis. James will be departing for a job as Senior Scientist at the chemical software company Schrodinger in New York City.

In memory of Brownwood, Baytown, Texas

For East Setauket, New York

And for all communities threatened by environmental destruction

When I have seen the hungry ocean gain

Advantage on the kingdom of the shore,

And the firm soil win of the watery main,

Increasing store with loss, and loss with store.

ACKNOWLEDGEMENTS

Thanks to Prof. Paulette Clancy, Clancy group alumni Dr. Jonathan Saathoff and Dr. Ananth Kaushik, and my mentees Jovana Andrejevic, James Raiford, Warren Crowell, Henry Herbol, and Blaire Sorensen.

Thanks to the Cornell Earth Energy IGERT, funded by the National Science Foundation's Integrative Graduate Education and Research Traineeship, award number 0966045.

Thanks to the National Science Foundation GK12, award number DGE-1045513, to Cornell's "Grass Roots: Advancing education in renewable energy and cleaner fuels through collaborative graduate fellow/teacher/grade-school student interactions."

CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Contents . . . vi
List of Tables . . . viii
List of Figures . . . x

1 Introduction **1**
1.1 The need for solution-processed solar materials . . . 1
1.2 Quantum dots . . . 3
1.3 Hybrid organic-inorganic perovskites . . . 7

2 Methods **11**
2.1 When to use all-atom models . . . 11
2.2 Choosing the right all-atom model . . . 12
2.3 Density Functional Theory . . . 12
2.4 Classical Force Fields . . . 15

3 The nucleation mechanism of PbS quantum dots in the Hines synthesis **19**
3.1 Methods . . . 21
3.2 Results . . . 29
3.2.1 Levels of *ab initio* theory . . . 29
3.2.2 Lead carboxylate hydrate . . . 30
3.2.3 Binding to TMS_2S . . . 33
3.2.4 The Zherebetskyy *et al.* initiation mechanism . . . 34
3.2.5 Alternative initiation mechanism . . . 35
3.2.6 Growth reaction mechanism . . . 37
3.3 Conclusions . . . 37

4 A simple molecular reactive force field for metalorganic synthesis **40**
4.0.1 Intermolecular force fields for reactive systems . . . 41
4.1 Methods . . . 43
4.1.1 Functional Form . . . 43
4.1.2 SMRFF Parameterization Method . . . 46
4.1.3 Parameter Optimization . . . 53
4.1.4 Evaluation of Error . . . 58
4.1.5 Implementation of a Reactive Dynamics Demonstration . 60
4.2 Results . . . 62
4.2.1 SMRFF Validation . . . 62
4.2.2 Reactive Dynamics . . . 68
4.3 Conclusions . . . 74

5 Rational Solvent Engineering for Hybrid Organic-Inorganic Perovskite (HOIP) Precursor Solutions **77**
5.0.1 Previous Theoretical Approaches . . . 78
5.1 Experimental . . . 81
5.1.1 *In silico* experiments . . . 83
5.1.2 *In vitro* experiments . . . 86
5.2 Results and Discussion . . . 87
5.2.1 Structures of Coordination Complexes . . . 87
5.2.2 Computation of ΔH_{solv} . . . 89
5.2.3 Experimental Solubilities . . . 91
5.2.4 Fast *ab initio* Prediction of Solubility . . . 96
5.2.5 Solvent mixtures . . . 99
5.2.6 Screening Alternative Solvents . . . 101
5.3 Conclusions . . . 102

6 Solvation modeling in highly exotic nonpolar solutions **106**
6.1 Methods . . . 109
6.1.1 Fugacity of the solid phase . . . 117
6.1.2 Polar PC-SAFT Equation of State . . . 118
6.1.3 Cubic Equations of State . . . 124
6.1.4 Parameterizations of the Equations of State . . . 126
6.1.5 Molecular Dynamics Simulations . . . 131
6.2 Solvation Results . . . 135
6.3 Azotosomes: Life in nonpolar solvents . . . 141

7 Conclusions and Future Work **161**

A A new algorithm for non-periodic Nudged Elastic Band **202**
A.1 Methods . . . 204
A.2 Results . . . 205

B Clancelot, a software package for computational chemistry **207**

LIST OF TABLES

4.1 Relevant abbreviations 52
4.2 A compilation of all interaction types and the potential function parameterized to handle each interaction. Note that electrostatic interactions were included for all interactions in addition to the appropriate potential shown in this Table. A key of all abbreviations is provided in Table 4.1. 57
4.3 Fitted Morse Potential Parameters 59
4.4 Lennard-Jones Potential Parameters 60
4.5 Charges 60
4.6 Root mean squared errors in energy for complex-to-dot and core-to-core interactions 66
4.7 Root mean squared errors in energy for overcoordinated systems 67
4.8 Root mean squared net forces on the complexes in optimized systems. Our target RMS net force, based on the "loose" optimization convergence criterion in Gaussian 09, was 20 kcal / mol / Angstrom [63]. 67
4.9 Mean absolute variations in atomic positions between dots optimized with SMRFF and DFT. 69

5.1 Experimental solubilities of lead halide ions for the pure solvents at room temperature 92
5.2 Common empirical properties of the solvents. We find that none of these properties are effective at ranking all the solvents considered here 97
5.3 Electronic properties of solvent molecules: note that the Mayer bond unsaturation and p-orbital occupation of the solvent's electronegative atom are predictive of the experimental solubility ranking, while the dipole magnitude, Loedwin atomic charge, and Mulliken atomic charge are not 98
5.4 Comparison of $\Delta H_{solv:Pb^{2+}}$, Mayer bond order (MBO) calculated at two different levels of theory, and experimental solubility of PbI_2 100
5.5 Proposed dative bonding solvents based on Mayer bond unsaturation of each solvent's double-bonded oxygen atom 102

6.1 The most common organic nitrogen species observed spectroscopically in Titan's upper atmosphere. Ranges refer to interpretations of the data *via* different ion chemistry models [123, p. 176]. 107
6.2 Solubility (mole fraction) of solid hydrocarbons in liquid methane at 110 K, liquid argon at 90 K and liquid krypton at 125 K. Experimental solubility ranges are taken from Szczepaniec-Cieciak [164]. *argon, **krypton. †Polar PC-SAFT, since COSMO-RS does not apply to noble gases. 114

6.3 Critical and melting properties of components used in this work [190, 191] . . . 127
6.4 Antoine equation parameters for components used in this work . 128
6.5 Pure component parameters for the non-polar terms in the PC-SAFT equation of state . . . 129
6.6 Pure component parameters specific to the Polar PC-SAFT equation of state . . . 129
6.7 Binary interaction parameters used in the SRK and Polar PC-SAFT framework . . . 130
6.8 Solubility in liquid methane using non-polar PC-SAFT (94 K, 1.5 bar), as a function of the binary interaction parameter k_{ij} . . . 130
6.9 Solubility in liquid methane using polar PC-SAFT (94 K, 1.5 bar), as a function of the binary interaction parameter k_{ij} . . . 131
6.10 Error in OPLS-optimized structures. Average absolute difference, for each species, between its OPLS structure and its structure *via ab initio* optimization with implicit solvent. . . . 133
6.11 Pairwise binding energies. OPLS and *ab initio* calculations produced similar results. . . . 134
6.12 Solubility (mole fraction) of nitriles and acetylene in pure methane, methane-nitrogen and methane-ethane-nitrogen mixtures using SRK, PR and COSMO-RS at 94 K. Mix. 1 = 90% methane, 10% nitrogen by mole. Mix. 2 = 50.1% methane, 41.3% ethane, 8.6% nitrogen by mole. **Indicates three-phase (vapor-liquid-solid) equilibrium. . . . 137
6.13 Solubility (mole fraction) of nitriles and acetylene in pure methane at 94 K and 1.5 bar, with and without including the polar term in the PC-SAFT framework. . . . 138
6.14 Free energies of solvation with respect to different thermodynamic states. The TI column should be considered most accurate, with the FEP columns interpretable as bounds for the sampling error. . . . 139
6.15 Polar nitrogen compounds found on Titan and their abundance in the upper atmosphere as measured by Cassini.[225, p. 167] . . 144
6.16 Error in OPLS-optimized structures. Average difference, for each species, between its OPLS structure and the structure obtained *via ab initio* optimization with Onsager's self-consistent reaction field (SCRF) model of an implicit solvent. . . . 154
6.17 Flexibility K_a of nitrile and amine azotosomes, and activation energy ΔE to remove a molecule from each azotosome. . . . 156
6.18 Gibbs free energy of decomposition: The net mechanical work required to remove a molecule from the membrane, within 20% uncertainty. These values are concentration-dependent. . . . 159

LIST OF FIGURES

1.1 Regions where the levelized cost of electricity (LCOE) from solar power is less than the average cost from the grid. Many more regions have solar power less expensive than grid power at peak hours. Source: Deutsche Bank Markets Research [1] 2

1.2 The fall in costs for the components of solar power installations in the USA. Source: Deutsche Bank Markets Research [3] 2

1.3 A lead sulfide quantum dot passivated with short organic ligands, with the sulfur mostly hidden in the core. Pb = brown, S = yellow, O = red, C = cyan, H = white. 4

1.4 Hybrid organic-inorganic perovskite, with iodine as the anion and methylammonium as the cation. Left: the unit cell. Right: the lattice. Pb = brown, I = pink, N = blue, C = cyan, H = white. . 7

3.1 PbS quantum dots passivated with COO^- only (left) and with OH^- as well (right). The left state has a great deal of steric unfavorability due to overlap between the oxygen atoms (red). 20

3.2 Energy surfaces which cause poor performance of NEB (top) and reaction coordinate methods (bottom). On these surfaces, the blue regions are low in energy and the dark arrow shows the path found by the algorithm. NEB overestimates the barrier if the true path is far from the geometric shortest distance. Reaction coordinate methods underestimate the barrier if it is orthogonal to the chosen reaction coordinate. 25

3.3 Nudged Elastic Band method running on the Zherebetskyy reaction mechanism at the B97-D3/TZVP level of DFT in implicit toluene solvent. The colors allow the different paths to be seen as the successive steps converge, in this case upward to a barrier around 250 kcal/mol. 26

3.4 Peak and endpoints of the NEB path for the Zherebetskyy reaction mechanism analyzed with pure and double-hybrid DFT. The difference between the peaks is large, 50 kcal/mol. 29

3.5 The isomers of a lead carboxylate hydrate monomer in nonpolar solvent. Left: no hydrogen bonds, middle: one hydrogen bond, right: two hydrogen bonds. The relative energies, in order, are +0, +0.7, and +3.1 kcal/mol, with the larger values indicating less stable states. Pb=brown,O=red,C=cyan,H=white. 31

3.6 A lead acetate hydrate dimer in nonpolar solvent. Note that both hydrogen bonding and the lead-oxygen bonds are fully satisfied in this configuration, with no competition between them. Pb=brown,O=red,C=cyan,H=white. 31

3.7 A lead acetate hydrate tetramer in nonpolar solvent. No additional hydrogen bonds are available in this structure as compared with the dimer, and steric hindrance is beginning to emerge. Pb=brown,O=red,C=cyan,H=white. 33
3.8 The Zherebetskyy initialization mechanism: reactants, intermediate, and products. The heavy atoms which form new bonds (Pb,S,Si,O) are circled throughout. 34
3.9 Our alternative initialization mechanism: reactants, intermediate, and products. The heavy atoms which form new bonds (Pb,S,Si,O) are circled throughout. 36

4.1 Key to the composition of our passivated PbS quantum dot systems delineated both by element (Pb, O, *etc.*) and in terms of the semiconducting PbS core and the surface complex. 41
4.2 Left to right, PbS quantum dots of sizes 1-6 (larger sizes become impractical in DFT). The size is defined by the number of S atoms (yellow), since the number of Pb atoms (brown) varies with the surface structure. The organic ligands on the surface have been omitted from the image, but were included in the optimizations to determine these structures. 44
4.3 Energy required to remove a given surface complex from the rest of the dot, for dot sizes 1-6, identified by the color key in the inset. Note that there is no simple relationship between dot size and binding energy, and that dots can have multiple distinct binding energies corresponding to removal of each of their possible surface complexes. In order to be useful, any model must reproduce this extraordinary variability. 47
4.4 *Ab initio* Pb-Pb dimerization at different levels of theory, compared with experiment. We selected HSE06/LANL2DZ, with an accurate minimum of 345 kJ/mol and slow long-range decay, avoiding the characteristic error of DFT in which interactions decay too quickly with distance [41]. 50
4.5 Comparison of interaction energies derived from several single point energy calculations using the M06L and HSE06 functionals. The plot shows the difference in energy between the two functionals as a function of different system geometries, each along a trajectory of removing one complex from the nanocrystal surface, color-coded for dot sizes 1-6 as defined in the inset. . . . 51
4.6 BSSE-corrected binding energy curves (solid) vs. BSSE-uncorrected curves (dashed). The uncorrected curves are always deeper because BSSE causes overestimation of binding energies. 52

4.7 The radial distribution functions between different atom types were used to provide initial parameters for their interactions. The location of the first peak of each curve (denoted by vertical lines) is the likely equilibrium distance for each pair. Left: interactions within dot cores; right: interactions between surfaces and cores. Color code as defined in the insets. 54
4.8 Complex-PbS surface interaction energies by dot size (from 1-6) as calculated using Gaussian 09 (shown as a blue band ±2 kcal/mol wide, see Supporting Information) and with the corresponding SMRFF predictions in red. Each curve corresponds to one complex (a ligand) being pulled away from the corresponding dot. 63
4.9 PbS-PbS core interactions calculated using Gaussian 09 (shown as a green band ±2 kcal/mol wide) and the corresponding SMRFF predictions in red. We define the core interaction as the binding energy between two halves of the core if it were divided into two charge-neutral pieces. Since larger cores can have multiple ways of dividing the core in half, we show one division of dot size 2 (top graph) and two divisions each of dot size 3 and 4 (second pair and bottom pair, respectively). 64
4.10 Gaussian 09-generated data (shown in blue) for several complex-to-nanocrystal interaction energies derived from overcoordinated structures, with the corresponding SMRFF fit in red, as a function of 359 distinct single-point calculations derived from several overcoordinated geometries. Results are shown for dot sizes from 1 (top plot) to 4 (bottom plot) because overcoordinated systems require more atoms than simply passivated systems, making overcoordinated dots of size 5 and 6 impractical in DFT. . . . 65
4.11 The most frequent interatomic distances for each interaction in our model, plotted against the equivalent distances in the DFT data set. The dashed line shows the ideal 1:1 ratio. The mean deviation from DFT distances is 6%. 68
4.12 Potentials of mean force for two metadynamics trajectories, 1 (blue) and 2 (orange). The inset images show the dots at the corresponding radii of gyration in trajectory 1. 70
4.13 Potential energy values plotted against the radius of gyration for trajectory 1 (left) and 2 (right). A smoothing filter has been applied to produce the dark blue line and extract an average trend in the raw (light blue) data. 71

4.14 Comparison of energy calculations performed *via* Molecular Dynamics (MD) using the metadynamics method and those performed using DFT for the first simulated trajectory. DFT calculations performed using the HSE06 functional are displayed in the top figure, and those using the M06L functional in the lower figure. 72
4.15 Comparison of energy calculations performed *via* molecular dynamics (MD) using the metadynamics method and those performed using DFT for the second simulated trajectory. Key as in Figure 12. 73
4.16 Schematic representation of the relative ordering of different levels of theory for all-atom simulation, including SMRFF. SMRFF is much faster than REAX or COMB with little sacrifice of accuracy. It is, however, transferable to fewer types of systems because of the limitations of its functional form, *e.g.*, no reactive angle dependent terms. 76

5.1 Pb^{2+} complex structures suggested by VSEPR theory for different numbers of complexing groups (3-8 from left to right) [108]. Our highly optimized complexes show many more ligands and less order than these regular polyhedra. Figure from Davidovich *et al.*, [108], used with authors' permission 80
5.2 The anti-common-ion effect observed for PbI_2 and KI, shown by the upward slope of the right side of the curve. This effect does not fully explain the effect of CH_3NH_3X on PbX_2 solutions, in which the solubility with CH_3NH_3X is *higher* than the solubility in the pure solvent, which would correspond to the rightmost point being higher than the leftmost point. Data from Lanford and Kiehl, 1941[110] 81
5.3 Comparison of DMSO (left) and acetone (right). They are analogous in structure *and* isoelectronic, differing in that DMSO is centered on a hypervalent sulfur atom, while acetone is centered on an octet-satisfying carbon atom. Colors: S=yellow, O=red, C=black, H=white 82
5.4 Comparison of DMF (left) and methacrolein (right) structures. They are analogous in structure but *not* isoelectronic, since methacrolein has an additional double bond which conjugates with its C=O double bond. Colors: O=red, N=blue, C=black, H=white 82
5.5 Comparison of two less common solvents for HOIP synthesis, GBL (left) and acetonitrile (right). They differ from the pattern set by DMSO and DMF in having either two oxygen atoms (GBL) or no oxygen atoms (acetonitrile). As will be shown, the electronegative atoms in the solvent play a critical role. Colors: O=red, N=blue, C=black, H=white. 83

5.6 Electron density predicted by our *ab initio* calculations surrounding a Pb^{2+}/acetone cluster with nine coordinated oxygen atoms, which we found to be the energetically preferred number of solvent neighbors. Color key: Pb = brown, O = red, C = cyan, H = white. The blue shading represents the electron density at the $0.005\ e^-/^3$ isosurface 88
5.7 Electron density predicted by our *ab initio* calculations surrounding a Pb^{2+}/DMSO cluster with eight coordinated oxygen atoms, the energetically preferred number of neighbors for this solvent. Color key: Pb = brown, S = yellow, O = red, C = cyan, H = white. The blue shading represents the electron density at the $0.005\ e^-/^3$ isosurface 89
5.8 *Ab initio* predictions of ΔH_{solv} for PbX_2, PbX^+, Pb^{2+}, and X^- complexes in pure solvents, with $CH_3NH_3^+$ as noted in the labels. Solvent key provided in the inset 90
5.9 Lead chloride perovskite "monomer" passivated by DMSO solvent. The Cl ions are stabilized by hydrogen bonds with $CH_3NH_3^+$. Color key: Pb=brown, Cl=orange, S=yellow, O=red, N=blue, C=cyan, H=white 92
5.10 *Ab initio* predictions of the structure of a (HOIP) perovskite "dimer" in a coordination complex with DMSO solvent. The methylammonium ions are clearly visible with their distinctive dark blue N atoms. Color key: Pb=brown, Cl=orange, S=yellow, O=red, N=blue, C=cyan, H=white 93
5.11 The energies of the sampled states of the Pb^{2+}/solvent complexes, with 50 samples for each solvent. The baseline for each curve is the lowest-energy state for that complex 93
5.12 Experimental solubility of PbX_2 *vs.* the *ab initio*-derived values of $\Delta H_{solv:Pb^{2+}}$ 95
5.13 Experimental solubility of PbX_2 *vs.* Mayer bond order of the oxygen atom in the solvent. The Mayer bond order follows the same trend as our calculated ΔH_{solv}, making it also an effective predictor of trends in solubility. The error bars reflect expected error in the solubilities and calculated Mayer bond orders. The color key for the halides is shown in the inset 99
5.14 $\Delta H_{cluster:Pb^{2+}}$ for clusters of DMSO:DMF, DMSO:acetone, and DMF:acetone. The color key for the solvent blends is shown in the inset 101

6.1 Schematic of the Polar and Perturbed Chain form of the Statistical Associating Fluid Theory (Polar PC-SAFT) 111

6.2 Vapor-liquid equilibrium (VLE) for light hydrocarbons and gases at conditions close to that on Titan. (a): methane-ethane [160]; (b): methane-propane [161]; (c): methane-nitrogen [162]; (d): methane-argon [163]; (e): ethane-argon [163]. (o): experimental data; red solid line (-): Polar PC-SAFT; black solid line (-): COSMO-RS; blue solid line (-): SRK. 112
6.3 Vapor-liquid equilibrium (VLE) of nitriles containing systems at high temperatures. (a): acrylonitrile-n-hexane [165]; (b): acetonitrile-benzene [166]; (c): acetonitrile-propanenitrile [167]; (d): acetonitrile-hydrogen cyanide [168]. (o): experimental data; red solid line (-): Polar PC-SAFT; black solid line (-): COSMO-RS; blue solid line (-): SRK. 115
6.4 Liposomes and azotosomes **A)** liposome in polar solvent. Polar heads are braced by non-polar lipid tails. **B)** azotosome in non-polar solvent. Non-polar tails are braced by polar nitrogen-rich heads. 145
6.5 Nitrogen-head positions in selected azotosomes. **A)** The initial grid, **B)** aminopentane (amorphous), **C)** pentanenitrile (hexagonal), **D)** acrylonitrile (close packed hexagonal). 149
6.6 Umbrella sampling of the azotosome decomposition process. The test molecule is incrementally withdrawn from the membrane in the z-direction. 150
6.7 Stretching the azotosomes. The slope of the fit line is proportional to the area modulus K_a. 151
6.8 Potential energy profile for the decomposition of acrylonitrile. The largest instantaneous energy barrier is the activation energy to decompose the azotosome. 152
6.9 Stretching a hexanenitrile azotosome and a hexane bilayer. The slope of the linear fit is proportional to the area modulus K_a. . . . 156
6.10 States of acrylonitrile. **A)** Azotosome. Interlocking nitrogen and hydrogen atoms reinforce the structure. **B)** Solid. Adjacent nitrogen atoms create some unfavorable repulsion. **C)** Micelle. Adjacent nitrogen atoms make this highly unfavorable. **D)** Azotosome vesicle of diameter 9 nm, the size of a small virus particle. . 158

A.1 NEB error with and without Procrustes rotation (see inset). Unfortunately, Procrustes rotation seems to make the error worse, not better. 205

CHAPTER 1

INTRODUCTION

1.1 The need for solution-processed solar materials

The economics of new solar power installations have improved dramatically over the past decade. The drop in cost, below the cost of grid electricity in several large markets (Figure 1.1), has created what Deutsche Bank has called a "gold rush" [1]. Ten US states have passed this point and thirty-six are expected to be past it by 2017 [1]. Once grid parity is surpassed, solar installations are limited only by the technical capacity of the grid to deal with intermittent sources of power. This capacity for modern grid systems appears to be at least 10%, the fraction of electricity generated by intermittent renewables in Hawaii in 2014 [2]. With improvements in energy storage technology, this capacity could theoretically be as high as 100% [3]. The pursuit of this unknown but apparently large slice of world energy markets can very accurately be called a gold rush.

The drop in the cost of solar power installations has been driven by a fall in panel prices, especially from China, which has led to increased volume and a consequent increase in economies of scale. The cost of a solar power installation is the sum of the costs of solar panels (including the solar cells and their housing), inverters, installation, wiring, and permits (Figure 1.2). A surprising reality is that the solar panels, while still the largest single cost, are no longer even close to the majority of the total cost of a solar installation.

The remaining "balance-of-system" costs (everything but the solar panel) can be reduced only gradually. The primary means of reducing these costs -

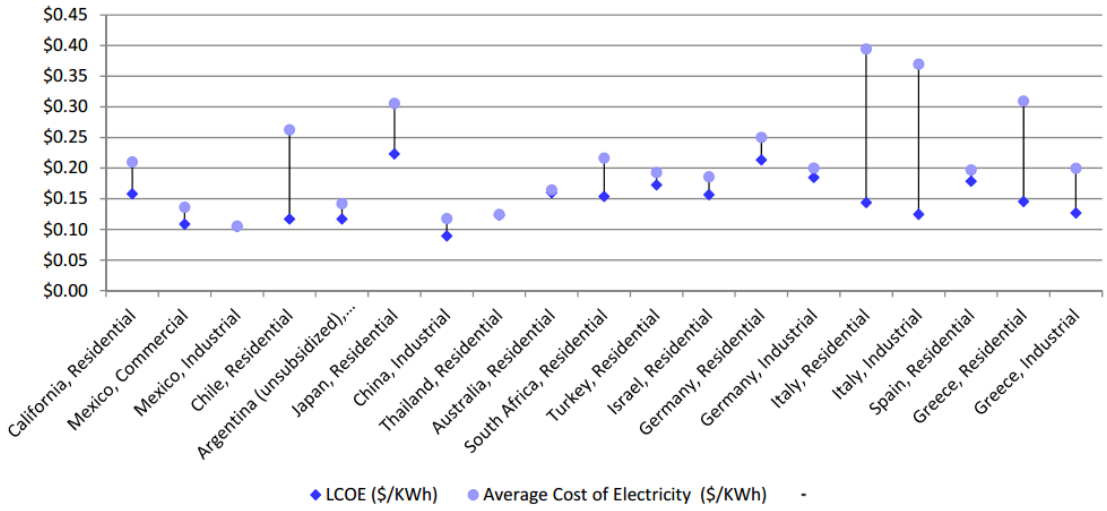


Figure 1.1: Regions where the levelized cost of electricity (LCOE) from solar power is less than the average cost from the grid. Many more regions have solar power less expensive than grid power at peak hours. Source: Deutsche Bank Markets Research [1]

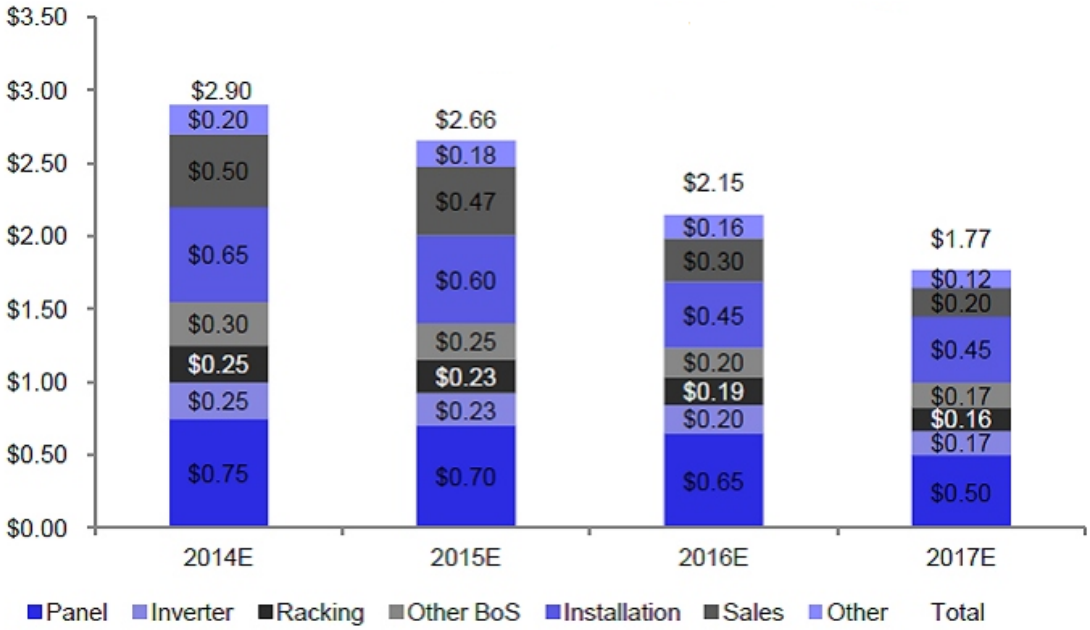


Figure 1.2: The fall in costs for the components of solar power installations in the USA. Source: Deutsche Bank Markets Research [3]

increasing experience with solar installations - is outside even the broadest definition of chemical engineering. However, balance-of-system costs can also be reduced by making solar cells which are inherently easier to deploy. Increases in efficiency and robustness, and reductions in size and weight, all tend toward this goal. The ultimate solar cells would be ones that could be added to roofing tiles, rolled out on flexible mylar sheets, or even applied as a liquid coating on site.

Current silicon solar cells are relatively stiff, heavy, and difficult to manufacture, which puts limits on how much progress is possible with regard to new designs. However, new materials which are the subject of avid research allow solar cells to be synthesized directly from organic solutions, which could bypass many of these limitations. The new solution-processed materials, semiconductor quantum dots and hybrid organic-inorganic perovskites, raise the possibility of even greater cost reductions and an even broader impact for solar power. Studies of the solution-processing of these materials will be the focus of this dissertation.

1.2 Quantum dots

A quantum dot is a semiconductor nanoparticle small enough that its electrons experience significant quantum confinement, causing size-dependent and therefore controllable electronic properties. Lead sulfide quantum dots in particular appear to be useful because they are relatively easy to synthesize at size distributions and band gaps which make them effective as solar cells. The study of lead sulfide quantum dots, less than 10 nm in size, is an active area of nan-

otechnology research: in addition to photovolatics, they have optical [4], energy storage [5], and even quantum computing applications [6]. Their chemistry allows engineered control over band gaps, crystal structures, lattice constants, and other properties previously considered inaccessible to change [5, 7, 8, 4, 9]. Quantum dots present promising prospects for the efficient harvesting of solar energy, in particular, tunable energy gaps and favorable photo-physics for multi-exciton generation and hot carrier collection. [5, 10] Quantum dot solar cell efficiencies have risen quickly over time, reaching 10.6% in 2016 [11].

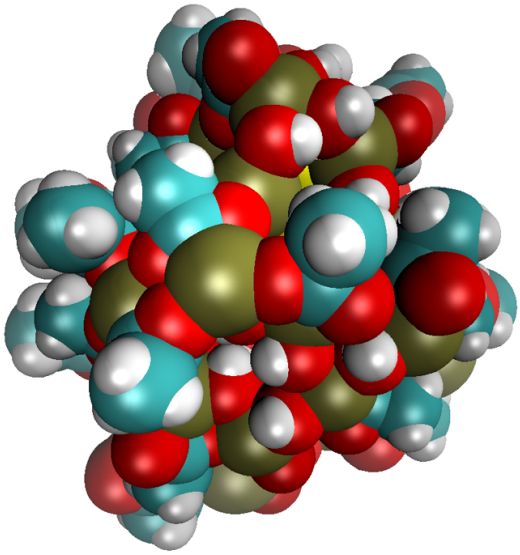


Figure 1.3: A lead sulfide quantum dot passivated with short organic ligands, with the sulfur mostly hidden in the core. Pb = brown, S = yellow, O = red, C = cyan, H = white.

Chemically, a lead sulfide quantum dot is composed of a PbS core, the structure of which approaches that of the bulk cubic lattice, with a less-ordered, passivated surface. The currently favored method for synthesizing PbS quantum dots is the Hines synthesis, in which lead oleate is the source of lead, and trimethyl silane sulfide (TMS_2S) is the source of sulfur [12, 4]. A large stoichiometric excess of lead oleate is used in the reaction, between three and ten

times the amount of sulfur. The sulfur source may be depleted by the reaction, but concentration of the Pb compounds and complexes remains significant throughout. Lead carboxylates, the family to which lead oleate belongs, have been synthesized for millenia in aqueous solutions (most famously the first artificial sweetener, "sugar of lead" [13]) but surprisingly, there is still uncertainty about their structure in non-aqueous solvent [4, 14, 15].

Lead sulfide quantum dot offer several advantages for electronic devices, most notably that they can be solution-processed at ambient temperature and potentially at low cost. However, achieving this promise of low-cost devices requires controlled manipulation of a large number of parameters: temperature, choice of suitable solvent(s), reaction time, surface ligands, and post-synthesis treatments [5, 7, 8]. The electronic properties of quantum dots are so sensitive that a high monodispersity of quantum dot sizes is required [5]. This large combinatorial set of options to control the manufacture of PbS quantum dots, and the tight constraints on size uniformity, points to the value of having a fundamental, theoretically based, understanding of the nucleation of lead sulfide quantum dots and, if possible, a link between their formation and the processing conditions by which they were produced. The processes which have been harnessed to create monodispersity thus far at the lab scale are not fully understood, making it difficult to maintain quality while scaling up. The prospects for large-scale PbS quantum dot manufacturing are thus limited by insufficient theoretical understanding of the factors that control dot nucleation and growth.

Only a small amount of theoretical work has been done to understand the synthesis of PbS quantum dots. The key paper is Zherebetskyy *et al.*, 2014 [4], which describes the putative reaction mechanisms behind the nucleation and

growth of the dots. Using Density Functional Theory (DFT), Zherebetskyy *et al.* calculated surface geometries for the quantum dots and the enthalpy of reaction for their mechanism and found it to be favorable [4]. However, the main focus of the work was experimental and the DFT was not pursued enough to calculate the kinetics of the mechanism [4]. More detailed DFT has been performed, but focused on the electrical properties of the finished dots, not on their processing [16, 17]. In terms of other levels of theory, attempts have been made to create fast models for the study of PbS quantum dots, using simplifications such as holding the core entirely rigid [18], or allowing the surface ligands to move but modeling them as charged spheres [19]. These simplifications allow the calculation of some properties of the finished dots, but not the study of the synthesis at any useful level of detail. A great deal more work needs to be done in simulating PbS quantum dots usefully with all atoms to increase the effectiveness of their solution processing.

Issues which need to be resolved are the nature of the surface passivation of the dots, the reason for the great excess of lead in the reactants for the Hines synthesis, and fundamentally, how the Hines synthesis works - its nucleation mechanism and its drivers of monodisperse quantum dot growth. Answering these questions via experiment has so far been impossible because the individual reactions cannot be resolved, but answering it via simulation requires performance which is only possible with a classical force field, of which none exists for this material. The creation of classical force fields has been important for the development of semiconductor materials ever since the Stillinger-Weber potential was developed to assist the field of silicon electronics. The development of a classical force field for PbS quantum dots should therefore be considered an important step in the progress of the study of this material.

1.3 Hybrid organic-inorganic perovskites

A hybrid organic-inorganic perovskite (HOIP) is a crystalline lattice of metal atoms (usually Pb, but conceivably Sn or Bi) held together by halide anions and organic cations (such as methylammonium or formamidinium). HOIP solar cell efficiencies have risen more quickly than those of any solar cell material ever, reaching 22% in 2016 [11]. Life-cycle analysis has shown that perovskite solar cells have the highest theoretical energy return on energy invested (EROEI) of any PV technology yet known, due to their low-energy solution-processing. [20] Perovskite solar cells also have a clear commercialization path, without needing to outcompete silicon first, *e.g.* as the upper layer in double-junction silicon-HOIP solar cells [21]. Such cells boost the practical efficiency of commercial solar panels from less than 20% to over 30% [21].

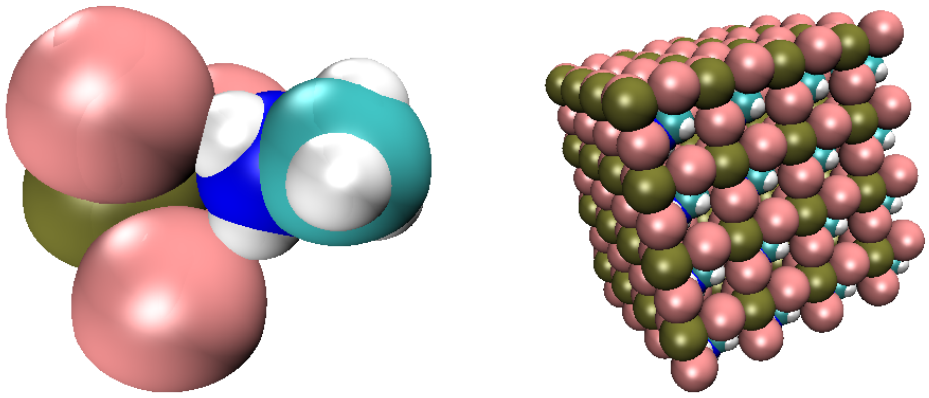


Figure 1.4: Hybrid organic-inorganic perovskite, with iodine as the anion and methylammonium as the cation. Left: the unit cell. Right: the lattice. Pb = brown, I = pink, N = blue, C = cyan, H = white.

However, our understanding of these HOIP precursor solutions remains poor. This may, in the future, cause the growth in HOIP efficiency to slow. Due to the immaturity of the field, the scale and consistency of synthesis processes

are problems for HOIP solar cells even more than for quantum dot solar cells [22, 23]. Since the perovskite films are about one third the thickness of a red blood cell, and since the precursor solutions are dewetting on the substrates, it is not surprising that pinholes can form during the synthesis process [24]. Apparently identical techniques yield cells with average efficiencies 4 to 10% lower than the expected highest efficiency [23]. Differences in humidity, air pressure, and the exact rate at which solvent is added to the system have been suggested as contributing factors to these large deviations. This inconsistency also manifests spatially, with smaller solar cells performing better than large ones: the best, 22% efficient cells have areas of order $0.1\,\mathrm{cm}^2$, while efficiencies fall to 15% for areas greater than $1\,\mathrm{cm}^2$ [25].

Stability with respect to time, light, and water damage are also important issues [26, 27], although they can be overstated. Perovskite films created with methylammonium cations tend to degrade via compositional changes (reversion to PbX_2), while films created with formamidinium tend to degrade via a crystalline phase change to a more stable but less photoactive state [26]. These concerns have somewhat been mollified by the mixture of methylammonium and formamidinium into single films, which has been observed to raise the barriers to both types of decay. However, the most important advances have been those in the consistency and crystallinity of the films. HOIP single crystals are very robust, but the crystalline fraction of currently-synthesized HOIP films is very low, perhaps as little as 30% [28]. Both stability and efficiency, therefore, we believe are related to improvements in the fundamental techniques of processing.

There are dozens of papers performing all-atom calculations on HOIP struc-

tures, but none on HOIP synthesis from solution. The reason for this appears to be that HOIP researchers are most interested in discovering the origins of perovskite efficiency and stability for the bulk films, and in seeking improvements through compositional changes. They have so far not shown an interest in changes to processing techniques. The following discoveries have all been made using Density Functional Theory all-atom studies:

- Without dispersion correction, DFT predicts HOIP lattice constants poorly. This implies that nonbonded van der Waals forces are structurally important within the crystal [29]
- DFT poorly predicts phase transitions within the HOIP film [29]
- DFT can be used to show how the cation moves within the unit cell [30, 31] and migrates between unit cells [32], possibly disordering the lattice
- DFT can be used to calculate the electron effective mass and hole effective mass in the HOIP lattice calculated, and the similarity between the two helps explain the electrical properties of HOIPs [29]
- When the HOIP central cation (methylammonium, formamidinium, cesium) is changed, the HOIP band gap does not respond as DFT predicts, even qualitatively [29]
- DFT predicts that there are ferromagnetic and antiferromagnetic regions within HOIPs, and that control of these properties could lead to cells with higher open-circuit voltage [33]
- DFT predicts that disorder and grain boundaries in HOIPs are relatively benign as regards electrical properties [29]
- DFT calculations of the enthalpy of formation of HOIPs showed that using mixed halides (Cl and Br along with I) makes the crystals more stable [29]

Because of this gap in research, our understanding of HOIP solution processing has lagged behind our understanding of the finished HOIP films. Critical questions, such as what makes HOIP solvents perform appropriately, have been left unanswered or answered incorrectly, leaving the composition of HOIP solutions largely a matter of guesswork. The Estroff group at Cornell has determined that replacing polar organic solvents with ionic liquids substantially increased the efficiency of their cells [34], but the best cells in the world, at 22% efficiency, have been made by using the complexation behavior of polar organic solvents [35]. It is unknown in the literature what form the coordination complexes of the Pb halide salts take in solution, with ill-founded expectations that they are sixfold-coordinated [36].

Similarly, there are ill-founded expectations that the solubility in a given solvent aligns with its relative polarity, and that the same goes for the strength of complexes with the solvents [37]. It is well known that methylammonium additives increase the solubility of the lead halides, but unknown how this occurs [38]. The effects of mixing solvents, an increasingly important technique for making the best solar cells [35], is likewise known only on an empirical basis for a few combinations of solvents. In the rush to create record efficiencies, these questions so necessary for further improvements and for eventual manufacturing and commercialization of these materials have been left unanswered. In Chapter 5 of this work, Rational Solvent Engineering for Hybrid Organic-Inorganic Perovskite (HOIP) Precursor Solutions, we will begin to answer them.

CHAPTER 2

METHODS

2.1 When to use all-atom models

All-atom models are the most detailed and most costly models used in computational chemistry. The fastest all-atom models, running for a week on a powerful parallel computer, will simulate a nanosecond of activity for less than a femtomole of material (100 million atoms). Therefore it is sensible to use all-atom methods carefully, and not apply them to problems which can be solved in any other way. In general, a good use of all-atom models involves phenomena too small and/or too rapid to observe experimentally, such as reaction, complexation, and binding mechanisms, or molecular packing geometries.

One common beginner's strategy in experimental chemistry is called "cook and look." An experimenter sets up a synthesis with an arbitrary combination of input variables, observes the outcome, and then tries a new combination until the desired outcome is reached. Computational chemists frequently offer to get their peers out of this jam. However, we are subject to a form of the same mistake: setting up a system, running a batch of expensive calculations, and repeating until some effect becomes obvious. A perfect example would be a long Molecular Dynamics run or a long relaxed energy surface scan simply intended to "see what happens."

Instead, the computational chemist must choose specific properties of the system in advance and design the calculations around them, such as ΔH or ΔG to see if a process will occur spontaneously, or E_a to estimate its rate, *etc*. The

properties of interest should be calculated at an inexpensive and approximate level of theory first, then at successively higher levels, so as to allow time for "sanity checks" and new ideas before committing to the more expensive main calculation. The selection of these lower and higher levels of theory is described below.

2.2 Choosing the right all-atom model

2.3 Density Functional Theory

Atomistic models of materials can be represented in two major ways: from first principles, or semi-empirically. The first-principles approach uses quantum mechanics, most popularly with the set of approximations known as Density Functional Theory (DFT) [39, 40, 41]. DFT produces accurate descriptions of many materials' properties (with some known deficiencies) and remains the "gold standard" of computational chemistry methods [41]. Within DFT, there are many possible approximations which can effect the efficiency and accuracy of the calculation.

There are two main types of DFT calculations, periodic and non-periodic. The periodic type, also called plane-wave DFT, represents the electron wavefunction as a weighted sum of periodic plane wave basis functions [40, 41]. In plane wave basis functions, the electron wavefunction $\phi(\vec{r})$ takes the form:

$$\phi(\vec{r}) = \sum_n c_n \exp\left(iG_n\vec{r}\right) \tag{2.1}$$

where $\vec{r}$ is the position vector, c_n are the weights, $i = \sqrt{-1}$, and G_n are vectors depending on the periodic box size and the wavenumber of this particular wave. By using a sufficient number of these functions with different wavenumbers, any periodic function can be represented, equivalent to the Fourier transform; in practice a finite cutoff is used which determines the efficiency and accuracy of the procedure [40]. The calculation of the weights is all that is necessary to determine the electron density and therefore the properties of the system. This calculation of weights requires a "functional," which provides the approximations necessary to calculate the energy of a given electron state, and a pseudopotential, a representation of the nucleus and core electrons of each atom [40]. Periodic DFT is very well-suited to three-dimensional repeating materials such as HOIPs, and so all DFT on HOIPs in the literature to date has been periodic [29]. However, periodic DFT is not nearly as useful for non-repeating systems such as occur during the solution-processing of a material. Periodic DFT can be applied to non-periodic clusters of atoms only by using a large simulation box and "padding" it with empty space, which is inefficient. For such systems, it is more sensible to use wavefunctions which are inherently non-periodic.

Non-periodic DFT generally represents the electron wavefunction with a set of Gaussian basis functions describing localized orbitals, which Stowasser and Hoffman describe as "a natural language for an *aufbau* of the complex reality of the molecules of the inorganic and organic world." [42] In Gaussian basis functions, the electron wavefunction $\phi(\vec{r})$ takes the form:

$$\phi(\vec{r}) = \sum_n c_n \exp\left(-\epsilon_n^2 \|\vec{r} - \vec{r}_n\|^2\right) \tag{2.2}$$

where $\vec{r}$ is the position vector, c_n are the weights, ϵ_n are the exponential con-

stants, and $\vec{r}_n$ are locations of the centers of the functions - unlike plane waves, Gaussian basis functions are localized in space. The positioning of the basis functions on each atom is determined by the basis set, with many sensible selections being available and thus many basis sets in wide use.

The selection of a sufficiently accurate basis is relatively straightforward because the basis functions cannot fit the underlying wavefunction better than a hypothetical perfect fit (the limit of an infinite basis), and therefore a better fit corresponds with lower DFT energy of a system. Therefore the basis which produces the lowest energy with a given functional is the most accurate for that functional, and the only questions are how much computation time and error are acceptable. The use of smaller basis sets followed by larger ones is a common technique to step through successively higher levels of theory [43, 44].

Comparison of functionals, by contrast, is difficult in the absence of experimental data because there is no way to tell whether a property predicted by a given functional is too high or too low. Therefore, it is best to find experimental data and perform calculations which can be compared to this data. If experimental comparisons are not feasible, it makes sense to use a benchmark study such as Goerigk 2011 [44] and choose the functional which performs best in the cases similar to one's own.

Metal complexes present particular challenges for DFT [45, 43]. First, metal reactions typically require a high level of *ab initio* theory, with large basis sets for isolated systems and high energy cutoffs for periodic systems. Second, because of their many accessible electrons and bonding states, metal complexes have a large number of possible structures, which makes them a challenge to optimize geometrically. Nonetheless, the study of metal complexes is tractable using a

combination of fast pure-DFT geometry optimization, hybrid or double-hybrid DFT for energy evaluation, and triple- or even quadruple- zeta basis sets with effective core potentials [45, 43, 44].

Unfortunately, the speed of DFT methods is very limited. The run time of these methods scales between N^2 and N^5 where N is the number of basis functions, with the higher-scaling methods being more accurate [46]. Therefore, while DFT calculations can be parallelized, generally they are applied to systems containing 200 atoms or fewer. This means that DFT is usable for only the smallest quantum dots and perovskite complexes, up to a few nm in diameter. Even then, DFT is best used in the role of calculating energy snapshots or minimizations and not for properties requiring tens of thousands of discrete samples. This makes determining a free energy landscape difficult with a DFT approach.

2.4 Classical Force Fields

Classical semi-empirical models of interatomic forces, or "force fields", provide the computational speed to enable free energy calculations on large systems, up to hundreds of thousands of atoms or more. Due to the use of optimizations such as cutoff radii and neighbor lists [47], the run time of these methods scales between N_a (linear) and $N_f \log(N_a)$ where N_a is the number of atoms. If properly parameterized and used within the scope of their parameterization, classical force fields can be as accurate as DFT methods or even more so (for example, a classical model parameterized against some experimental properties will fit similar properties more accurately than all but the most costly DFT). However,

classical force fields do not fit the underlying physical world as closely as DFT, and thus are more vulnerable to misuse. Like all empirical models, classical force fields are not inherently transferable, and the reparameterization process can be quite arduous. A great many types of force fields exist in the literature, each with its own functional form [48, 49, 50, 51, 52, 53] (see especially the 2013 review by Sinott and van Duin, [54]). Because of the reactive nature of our chosen systems, we will only consider force fields which allow the breaking and forming of bonds.

The simplest force fields which might be considered as representing the breaking and forming of bonds are those designed to capture specific bonding arrangements in solid phase materials through a suitably constructed three-body term: Stillinger-Weber [48], Tersoff [49], and the Modified Embedded Atom Model (MEAM) [50]. These potentials have a long and impressive record of application, with Stillinger-Weber, Tersoff and MEAM designed to model semiconductors which bond in a diamond cubic arrangement. Similarly, the Embedded Atom Model (EAM) [51] is designed to represent the delocalized electron density characteristic of metals. In contrast to many other reactive potentials, these models are robust enough to allow timesteps of one femtosecond or higher to be used in Molecular Dynamics simulations, making them particularly efficient in terms of computer resources. However, this class of potentials does not effectively incorporate long-range interactions between non-bonded atoms beyond the first or second nearest neighbor shells [54]. The Tersoff-based AIREBO (Adaptive Intermolecular Reactive Empirical Bond-Order) [52, 55, 54] model attempts to address this concern by incorporating non-bonded terms into a Tersoff-style local bonding model. However, the AIREBO formulation is tuned specifically for hydrocarbon molecules, losing applicability to heterogeneous

systems in the process [54].

The Bond Order Potential (BOP) attempts to solve this transferability problem by building from a sound physical foundation.[56] BOP is derived from tight-binding quantum mechanics with a one-electron approximation, and many of its parameters directly reflect the electronic structure of the atoms. It has been used for several binary semiconductor systems, such as CdTe and GaAs [56], with somewhat better and more transferable results than Tersoff or REBO potentials [56]. However, due to its complexity, BOP has only ever been parameterized for one ternary system, Cd-Zn-Te [57], and never for five interacting elements, such as we will require to study the PbS and organic ligands in PbS quantum dots or the Pb halide and organic cations in HOIPs.

The most popular reactive potentials today are the "Charge-Optimized Many Body" (COMB) model [54] and the "Reactive Force Field" (REAX) [53]. Both potentials employ a semi-empirical, semi-physical functional form, with separate terms for all the types of bonded and non-bonded interactions which atoms are known to undergo. Both of these potentials allow charges to vary during bonding, making their long-range Coulombic interactions more accurate than those of any other potential. COMB has been studied for heterogeneous metal-organic systems, though not typically for systems containing many atom types. [54] REAX is being applied to a variety of applications across many types of materials, from hydrocarbons to semiconductors [54]. Of necessity, given their formulation, both COMB and REAX models require hundreds of parameters, and the parameter-fitting process is neither straightforward nor fast. An important consideration, therefore, is how to improve the process of parameterizing reactive force fields. But before we parameterize a force field, we must under-

stand the fundamentals of the underlying mechanism.

CHAPTER 3

THE NUCLEATION MECHANISM OF PbS QUANTUM DOTS IN THE HINES SYNTHESIS

Zherebetskyy *et al.* greatly enhanced our understanding of the role of lead carboxylate in the PbS synthesis with their groundbreaking 2014 paper entitled: "Hydroxylation of the surface of PbS nanocrystals passivated with oleic acid." [4] This paper uncovered three important facts: First, the lead oleate used in the Hines synthesis is, in fact, a hydrate, with the water molecule so tightly bound to the lead atom that it had not previously been detected. Second, this newly discovered water plays an active role in the PbS nanoparticle synthesis. Third, the resulting nanoparticles contain hydroxide ligands, in addition to the oleate ligands that had been known previously.

As surprising as these revelations may have seemed initially, our viewpoint is that they resolved several outstanding questions in this area. The oleate ligand is terminated by a carboxylate group, a somewhat bulky group due to its two oxygen atoms. Steric hindrance prevents a surface composed of lead ions from bonding with more than one carboxylate group each. However, one carboxylate group (oxidation state -1) is not sufficient to fully passivate a lead ion (oxidation state +2). Furthermore, we know that PbS quantum dots have substantial numbers of excess lead atoms on their surfaces [58]. Introduction of the suggestion that hydroxide ions are present at the quantum dot surface explains all these observations: each lead atom can have two anionic ligands, oleate and hydroxide, without violating steric hindrance, as shown in Figure 3.1. Thus, excess lead atoms can passivate the whole surface and be passivated themselves

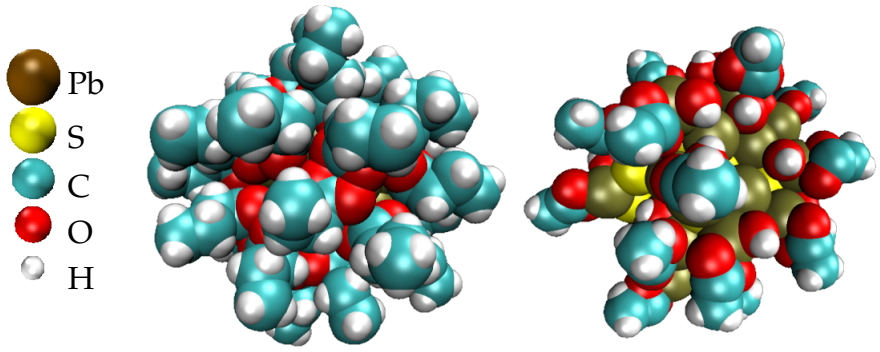


Figure 3.1: PbS quantum dots passivated with COO^- only (left) and with OH^- as well (right). The left state has a great deal of steric unfavorability due to overlap between the oxygen atoms (red).

at a +2 oxidation state with no contradiction.

The mechanism proposed by Zherebetskyy *et al.* is a substitution reaction in which the central sulfur atom of TMS_2S is exchanged with the central oxygen atom of lead oleate hydrate. The two carboxylate groups then join and form (Oleyl-CO)$_2$O-H_2O, leaving PbS, as described by the equation:

$$(Oleyl - COO)_2Pb - H_2O + (CH_3)_3Si - S - Si(CH_3)_3 \rightarrow$$
$$(CH_3)_3Si - O - Si(CH_3)_3 + (Oleyl - CO)_2O - H_2O + PbS \tag{3.1}$$

They propose that (Oleyl-CO)$_2$O-H_2O later rearranges into a hydrogen-bonded dimer of oleic acid.

Zherebetskyy *et al.* tested their mechanism using Density Functional Theory (DFT) studies, but only on the equilibrium states to determine ΔH_{rxn}. In this chapter, we shall use DFT calculations of transition states to show that the reaction mechanism proposed in Zherebetskyy *et al.*'s paper does not represent the lowest energy path, particularly in the nucleation steps. We will demonstrate

that their proposed mechanism would encounter unfavorable energy barriers and we will propose an alternative mechanism, with lower-energy intermediates, which will round out the theory presented in Zherebetskyy *et al.*'s illuminating paper.

3.1 Methods

In this study, we employed both periodic and non-periodic DFT. For our periodic, plane-wave calculations, we followed Zherebetskyy *et al.* in selecting the Projector Augmented Wave (PAW) pseudopotentials [4]. We used the Perdew-Burke-Ernzerhof (PBE) functional [59], a more commonly used and more recent functional in the family of the Perdew-Wang-1991 (PW91) functional used by Zherebetskyy *et al.* [60]. We performed our plane-wave calculations using the Quantum Espresso [61] DFT package with a kinetic energy cutoff of 30.0 Ry (400 eV). For geometry optimization, we used a force convergence threshold of 0.01 eV/ as in Zherebetskyy, and we added an energy convergence threshold of 0.0001 Har (0.0027 eV) for additional convergence.

We used these periodic calculations to validate our second approach, non-periodic DFT. These localized DFT studies, which constitute the bulk of our calculations on the Hines mechanism, employed the DFT package Orca [62] with some initial work in Gaussian 09 [63].

For the initial calculations in Gaussian 09, we used the Correlation-Consistent Polarizable Valence (*) Zeta (cc-pV*Z) family, where (*) describes the level of multiple basis functions used, which correspond to different valence atomic orbitals and are denoted as double-zeta, triple-zeta, *etc.* [64]. For this sys-

tem of atoms, we found it necessary to select the triple-zeta cc-pVTZ basis set. For the functional, we selected the Heyd-Scuseria-Ernzerhof method (HSE06), a range-separated DFT hybrid [65]. The popular HSE06 functional, based on the Perdew-Burke-Ernzerhof (PBE) family, [59] is a good choice for metal-containing systems [66]. It has the advantage of a long-range correction which makes it more accurate for systems with longer ranged non-bonded interactions [67]. To compare with the calculations in Zherebetskyy *et al.*'s paper, we also performed some calculations in Gaussian 09 with the PW91 non-hybrid functional. For atomic orbital calculations with heavy elements, such as Pb, it is also helpful to replace the inner electrons with an Effective Core Potential (ECP). The principal core potentials belong to three families: the Compact Effective Potentials [68], the Los Alamos National Laboratory Double-Zeta (LANL2DZ) ECP [69], and the Stuttgart-Dresden-Bonn (SDB) ECP [70]. All of these core potentials have been benchmarked and proven to be effective for organic molecules containing a metal atom [71], while LANL2DZ and SDB are preferred for a system like ours with polar lead-oxygen bonds [72]. Although LANL2DZ and SDB are parameterized differently - LANL2DZ with quantum calculations on isolated atoms, whereas SDB is semi-empirical - they give similar results for the Pb systems we studied. The SDB potentials are generally acknowledged to be more reliable because they include more of the core electrons [72]. For the atoms with core potentials, we used each core potential's Gaussian09 default basis for the valence atoms. To represent the solvent, toluene, we used the Polarizable Continuum Method (PCM) with the integral equation formalism variant (IEFPCM) [73].

For the final calculations in Orca, we used the fast B97-D3 functional for geometry optimizations, and the double-hybrid PWPB95-D3 functional for the final single-point energy calculations, both recommended in Grimme's DFT

benchmark study [44]. We used the RIJK approximation for all integrals [62]. We performed these calculations with Truhlar's def2-*VP family of basis sets, which is more accurate than the cc-pV*Z family, at the cost of not being designed to calculate the basis set limit for properties by extrapolation[74, 64]. Since we did not use such extrapolation, Truhlar's basis sets were more efficient for our application than the cc-pV*Z basis sets. We used the Stuttgart ECPs for lead atoms as in our Gaussian 09 calculations [72]. We represented the solvent using the more modern 'Solvation Model based on Density' (SMD)[75]. All of the final presented energies in this chapter were calculated at this level of theory, RI-PWPB95-D3/def2-TZVP in SMD-modeled toluene solvent.

To reduce the computational resource demands of the system, we removed the carbon tails of the oleate ions after the first C-C single bond, turning them into ethanoates. Past the non-conductive C-C single bond, the electronic states of the reactive site are unlikely to be affected by the presence of the remaining carbon chain, so it can be safely truncated for most purposes [76]. The long carbon chain is necessary to improve the solubility in practice, but this is not an issue in our DFT calculations.

Binding energies and reaction energies can be calculated using only the energies of the start and end states, optimized to their local energy minima. However, in order to determine the activation energy E_a of each step of the reaction, we must find all the transition states that are involved. Transition states are by no means simple to determine because a transition state is inherently not the minimum, and therefore geometry optimization leads away from it, not toward it. Therefore a whole class of methods has been developed to deal with locating transition states. The leading method for generating transition states is the

Nudged Elastic Band (NEB) [77] method. In this method, multiple geometry optimizations between the reactants and products are performed in parallel, each optimization being coupled to the adjacent ones by harmonic constraints which keep the structures geometrically linked [77]. Nudged Elastic Band has the convenient property of being *variational* - that is, given a sufficient discretization of the path, it will never find an energy path lower than reality [77]. This property is shared by other path-based algorithms such as Steered Molecular Dynamics and Forward Flux Sampling, but not by higher-order optimization algorithms such as Synchronous-Transition Quasi-Newton method, which can converge to the wrong saddle point in the energy surface [43, 78]. Similarly, the variational property is not shared by reaction coordinate methods such as Thermodynamic Integration [79] and Adaptive Biasing Force [80], which can skip certain energy barriers.

The tradeoff is that the Nudged Elastic Band algorithm does not exhaustively explore the configurational space because the harmonic constraints weight it toward the geometrically shortest path between the given start and end points. The start and end geometries are completely fixed and therefore must be correct for NEB to produce a meaningful outcome. By contrast, methods such as Thermodynamic Integration only restrain the system along the single dimension of the pre-determined reaction coordinate, sampling all other dimensions freely. However, the choice of this reaction coordinate for complicated systems can be even more difficult than the choice of start and end states for NEB: if the reaction coordinate is chosen incorrectly, the true barrier may be missed entirely, and the result can be either higher or lower than the true value. Examples of energy surfaces which cause problems for each algorithm are shown in Figure 3.2.

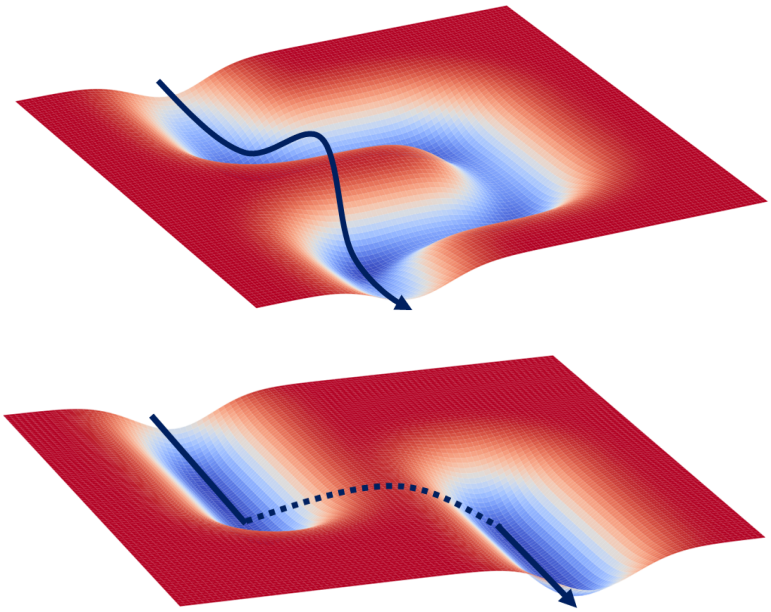


Figure 3.2: Energy surfaces which cause poor performance of NEB (top) and reaction coordinate methods (bottom). On these surfaces, the blue regions are low in energy and the dark arrow shows the path found by the algorithm. NEB overestimates the barrier if the true path is far from the geometric shortest distance. Reaction coordinate methods underestimate the barrier if it is orthogonal to the chosen reaction coordinate.

In summary, NEB is the best choice for metalorganic-complex reactions such as this one, where good start and end geometries are available but the choice of a reaction coordinate would be difficult. Because Orca and Gaussian09 do not support NEB natively, we created a custom NEB driver for both programs (Figure 3.3).

To confirm the transition states generated *via* NEB, we performed geometry optimizations from either side of each peak and confirmed that these led back to the expected start and end points [77]. To ensure that these optimizations would follow the energy surface to the local optimum from each point and not skip any energy barriers, we performed these optimizations with a dynamic trust radius

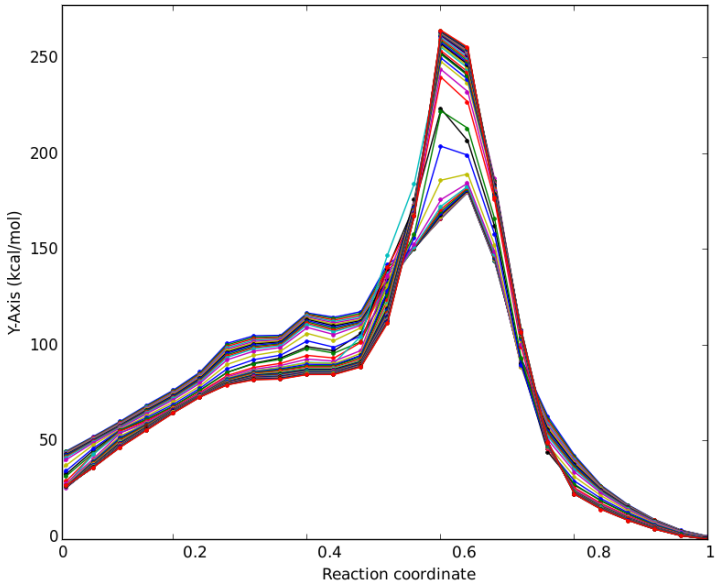


Figure 3.3: Nudged Elastic Band method running on the Zherebetskyy reaction mechanism at the B97-D3/TZVP level of DFT in implicit toluene solvent. The colors allow the different paths to be seen as the successive steps converge, in this case upward to a barrier around 250 kcal/mol.

starting at just 0.03 Angstroms (10% of the Orca default [62]).

In non-periodic DFT, the basis of calculation for the wavefunction is a function of the atomic positions, because the basis functions follow the atoms. This means that binding energies and geometries calculated with small basis sets can have significant error due to the overlap of basis sets on adjacent atoms, which produce spurious energies due to better fitting of the wavefunction at the locations where more basis functions are present [81]. In the Gaussian 09 single-point calculations with LANL2DZ basis sets (a small basis set), we corrected for the basis set superposition error (BSSE) [81] using "ghost atoms." Ghost atoms are atoms in a non-periodic DFT calculation which have basis functions but no nuclei or electrons [63] - thus they affect how well the wavefunction is repre-

sented, but do not otherwise contribute to the energy. The construction of a system with a subset of its real atoms changed to ghost atoms allows us to estimate how much the energy of the system is lowered by the extra basis functions from this subset of atoms. The equations used are:

$$E_{binding} = E_{interact} + E_{deform} \tag{3.2}$$

Which breaks the binding energy into interaction energy (negative) indicating how much the two parts attract each other and deformation energy (positive) indicating how much energy it takes to change the geometry of the parts such that they can bind together. The interaction energy is broken down into:

$$E_{interact} = E(AB)^{AB}_{AB} - E(A)^{AB}_{AB} - E(B)^{AB}_{AB} \tag{3.3}$$

where $E(AB)^{AB}_{AB}$ is the energy of the full system, $E(A)^{AB}_{AB}$ is the energy of part A evaluated using the basis functions of the full system, and $E(B)^{AB}_{AB}$ is the energy of part B evaluated using the basis functions of the full system. The basis functions of the full system (indicated with superscript AB) are added to the parts A and B using ghost atoms. All of these evaluations are performed in the optimized geometry of the full system (subscript AB). Since this is not the geometry that either part would have in isolation, we add the deformation energies for parts A and B:

$$E_{deform} = E_{deform,A} + E_{deform,B} \tag{3.4}$$

$$E_{deform,A} = E(A)^{A}_{AB} - E(A)^{A}_{A} \tag{3.5}$$

$$E_{deform,B} = E(B)^{B}_{AB} - E(B)^{B}_{B} \tag{3.6}$$

where $E(A)^{A}_{AB}$ is the energy of A evaluated using its own basis functions in the geometry of the full system, and $E(A)^{A}_{A}$ is the energy of A evaluated using its own basis functions in its own optimized geometry. The calculation for $E_{deform,B}$ is analagous to that for $E_{deform,A}$. The sum of these terms is the full deformation energy, which when added to $E_{interact}$ gives the BSSE-corrected binding energy between A and B.

For the final Orca calculations, we used a different approach. In general, we used basis sets of triple zeta size to reduce the ill-fitting of the wavefunction which causes BSSE. When we did use double-zeta basis sets, we used them with the Gaussian Counterpoise Correction (GCP) method, which counteracts BSSE using empirical repulsion terms between the atoms [82]. GCP is fast and simple enough that it can be applied even during geometry optimizations. This allows even small basis sets to produce a reasonable guess at the geometry. We performed the final optimizations with the triple-zeta def2-TZVP basis set which is large and accurate enough not to benefit from the GCP correction. This is particularly true for our system since Pb is not yet included in the GCP empirical parameter set, making it less accurate for Pb than for other elements at the time of this writing [82].

3.2 Results

3.2.1 Levels of *ab initio* theory

Comparing periodic and non-periodic DFT, we optimized the geometries of lead carboxylate hydrate monomers. For both periodic and non-periodic pure DFT calculations, the result was a Pb-O distance (for the H_2O oxygen) of 0.28 nm. This supports the comparison of our work with that of Zherebetskyy: using sufficiently large basis sets, such as the triple-zeta def2-TZVP basis set we used, we see no significant difference between periodic and non-periodic results at the same level of theory.

As part of analyzing our reaction paths, we also compared the results of the pure DFT B97-D3 functional with those of the more costly double-hybrid DFT PWPB95 functional, as shown in Figure 3.4.

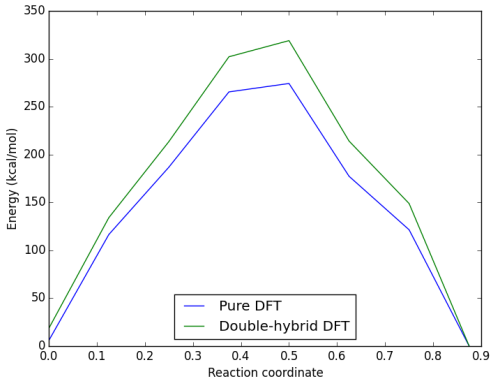


Figure 3.4: Peak and endpoints of the NEB path for the Zherebetskyy reaction mechanism analyzed with pure and double-hybrid DFT. The difference between the peaks is large, 50 kcal/mol.

As expected, the two energy curves have the same shape but different mag-

nitude. This means that the pure DFT functional has maxima and minima at approximately the same geometries as the double-hybrid functional, allowing it to be sufficiently accurate for geometry optimization, but that the final energy evaluation should be done at the higher level of theory, as per our procedure.

3.2.2 Lead carboxylate hydrate

Zherebetskyy *et al.'s* [4] first key insight was that lead carboxylate in the Hines synthesis appears in the form of a hydrate. Our calculations fully support this view. The binding energy of the water molecule to the lead carboxylate molecule in toluene is large, 20 kcal/mol, and the resulting Pb-O distance is 0.28 nm, only 0.03 nm greater than the average bond length of the lead carboxylate bonds, 0.25 nm. Therefore, we support the assertion that latent water will remain in lead carboxylate compounds in the hydrate state unless extreme attempts are made to remove it using heating or vacuum [4].

We find that a lead carboxylate hydrate complex in a nonpolar solvent has three isomers which are close in energy, differing in how many internal hydrogen bonds are formed between the water molecule and the carboxylate ions (Figure 3.5). Hydrogen bonding is frustrated in a lead carboxylate hydrate molecule because, in order to form hydrogen bonds, the carboxylate ions have to weaken their bonds to the Pb^{2+} ion. Despite the general favorability of hydrogen bonding, the isomer without hydrogen bonds has the lowest energy because it allows stronger bonds between Pb and the carboxylate oxygen atoms.

As pointed out by Zherebetskyy *et al.*, hydrogen-bonding of the reactants is a strong driver of complexation in nonpolar solvents [4]. However, what Zhere-

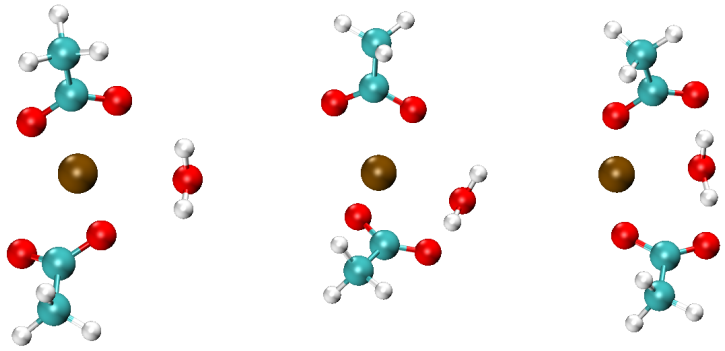


Figure 3.5: The isomers of a lead carboxylate hydrate monomer in nonpolar solvent. Left: no hydrogen bonds, middle: one hydrogen bond, right: two hydrogen bonds. The relative energies, in order, are +0, +0.7, and +3.1 kcal/mol, with the larger values indicating less stable states. Pb=brown,O=red,C=cyan,H=white.

betskyy did not take into account is the fact that the same might hold true for the lead carboxylate hydrate itself. The frustration of the internal hydrogen bonding in lead carboxylate hydrogen can be resolved if one molecule of it hydrogen bonds with another, producing the desired hydrogen bonds without other unfavorable changes to the geometry (Figure 3.6).

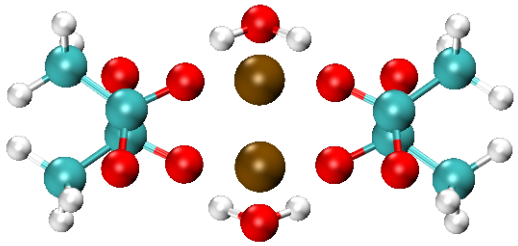


Figure 3.6: A lead acetate hydrate dimer in nonpolar solvent. Note that both hydrogen bonding and the lead-oxygen bonds are fully satisfied in this configuration, with no competition between them. Pb=brown,O=red,C=cyan,H=white.

This gives lead carboxylate hydrates in nonpolar solvents a strong tendency to dimerize, in a process that we found has no activation energy barrier and a

significant binding energy of -8.5 ± 1 kcal/mol/monomer. This makes sense because dimerization allows all of the hydrogen-bonding of the system to be fully satisfied, while still allowing all the oxygen atoms to bind to the lead. Furthermore, this configuration results in the carboxylate carbon chains pointing in opposite directions, producing no steric hindrance regardless of their lengths. As a result, we believe that the monomer does *not* play a significant role in the system. Our discovery implies that the dimer form of lead carboxylate hydrate will be the dominant form in nonpolar solvents, including in the Hines synthesis.

We also tested the likelihood of tetramerization, but found it to have a binding energy of only -3.6 ± 1 kcal/mol/monomer, compared to -8.5 ± 1 kcal/mol/monomer for the dimer. Considering the entropic contribution of remaining in the dimer state, and the steric hindrance which the tetramer will produce for longer carboxylate carbons chains (Figure 3.7), this low binding energy makes the lead carboxylate hydrate tetramer unlikely to be a common structure. This is understandable because, having satisfied both lead-oxygen bonding and hydrogen bonding *via* dimerization, lead carboxylate hydrate has no remaining driver for tetramerization except for electrostatic and van der Waals interactions. Therefore, we can confirm that most lead carboxylate in solution in a nonpolar solvent, as in the Hines synthesis for nanoparticles, will be found in the dimer state.

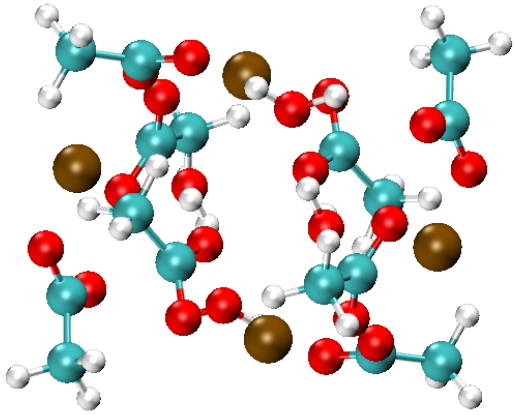


Figure 3.7: A lead acetate hydrate tetramer in nonpolar solvent. No additional hydrogen bonds are available in this structure as compared with the dimer, and steric hindrance is beginning to emerge. Pb=brown,O=red,C=cyan,H=white.

3.2.3 Binding to TMS_2S

The next step in the reaction involves the binding of the lead carboxylate hydrate with the sulfur source TMS_2S. This binding, driven by simple electrostatic interactions between Pb and S, precedes any reaction, but is important because it reduces the concentration of free TMS_2S. For a lead carboxylate hydrate monomer, the binding energy is -15.8 ±1 kcal/mol, and for the dimer, -23.9 ±1 kcal/mol. These large interaction energies shift the expected equilibrium so far toward the bound state that no free TMS_2S can be expected in the solution once a large stoichiometric excess of lead carboxylate is added. This one-way nature of the system has implications for the nucleation and later growth stages of the reaction, as will be discussed in later sections.

3.2.4 The Zherebetskyy *et al.* initiation mechanism

In the Zherebetskyy initiation mechanism, the water molecule is broken apart and its oxygen atom is exchanged with the sulfur atom of TMS_2S. Our NEB trajectory for this reaction is shown in Figure 3.8.

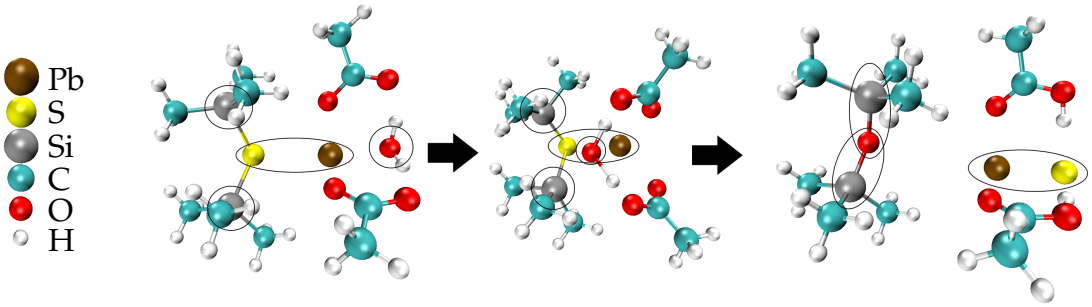


Figure 3.8: The Zherebetskyy initialization mechanism: reactants, intermediate, and products. The heavy atoms which form new bonds (Pb,S,Si,O) are circled throughout.

From analysis of the start and end states, we found a ΔH_{rxn} for Zherebetskyy's mechanism of -18.4 ±1 kcal/mol, which is favorable. However, via NEB we calculated an activation energy E_a of 300 ±1 kcal/mol, which is highly unfavorable. The energy barrier arises largely from the fact that the sulfur and oxygen atoms must switch places to transform TMS_2S to TMS_2O, with both TMS^+ ions leaving the sulfur and joining the oxygen. Even though this transformation has a favorable ΔH_{rxn}, it involves the creation of a high-energy intermediate in which the TMS^+ ions, S^{2-} ion, and O^{2-} ion form an unstable cluster (the center image in Figure 3.8). This produces a very high E_a.

We also observe that as the water molecule dissociates in the presence of TMS^- ions, it strongly prefers to form (TMS)OH rather than TMS_2O. This happens even though the O-H bonds in the water molecule are allowed to, and in

fact do, break one at a time during the reaction. The formation of TMS_2O is apparently not likely to occur. A further problem for this reaction mechanism is that it relies on non-dimerized lead carboxylate. As we have shown, lead carboxylate monomers will be rare, making a mechanism which relies on them less likely.

Finally, the formation of (Oleyl-CO)$_2$O-H_2O (oleic anhydride), as described in the Zherebetskyy mechanism, does not appear likely. The formation of oleic acid is not a limiting step in the reaction; it is formed by a simple proton transfer from H_2O to oleic acid, while the formation of oleic anhydride would require further reactions. Furthermore, the acid is experimentally known to be more stable than the anhydride+water complex by 13.5 kcal/mol [83]. For these reasons, we believe that the reaction pathway does not include (Oleyl-CO)$_2$O-H_2O.

In answer to these objections, we present an alternative mechanism which avoids all the concerns listed above, and which we will demonstrate offers a more kinetically favored route than that originally proposed by Zherebetskyy *et al.* [4].

3.2.5 Alternative initiation mechanism

Our proposed initiation mechanism is given by:

$$[(Oleyl-COO)_2Pb-H_2O]_2 + (CH_3)_3Si-S-Si(CH_3)_3 \rightarrow$$
$$[PbS][(Oleyl-COO)_2Pb-(H_2O)_2] + 2(CH_3)_3Si-(Oleyl-COO) \qquad (3.7)$$

In other words, the silicon of TMS_2S reacts with the oxygen of a carboxylate group to form R-COO-TMS, allowing the Pb to react with the S from TMS_2S. This hypothesis arises naturally from the geometry of the complex formed by our discovered lead carboxylate dimer and the TMS_2S. In this complex, the carboxylate ions are very close to the silicon atoms of TMS_2S. This suggests an idea which turns out to be true: the easiest reaction route is for the TMS silicon to react with a carboxylate oxygen, not with the oxygen atom in the water molecule.

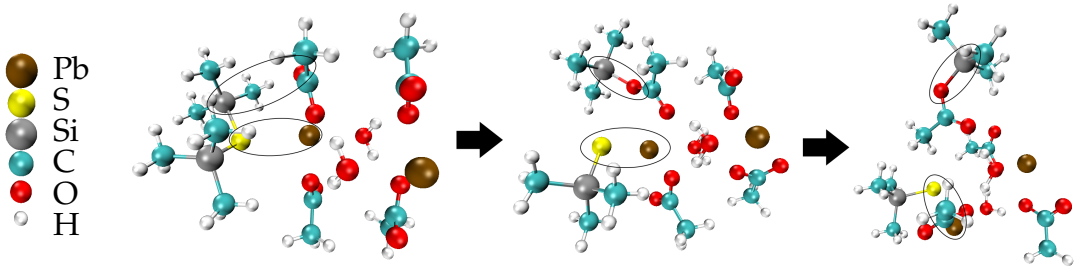


Figure 3.9: Our alternative initialization mechanism: reactants, intermediate, and products. The heavy atoms which form new bonds (Pb,S,Si,O) are circled throughout.

From analysis of the start and end states, we found a ΔH_{rxn} for our mechanism of -4.0 ±1 kcal/mol, which is favorable. Using NEB, we calculated an activation energy E_a of 65 kcal/mol, which is a fraction of the large value we obtained for the Zherebetskyy mechanism. The ΔH_{rxn} of this pathway, -4.0 ±1 kcal/mol, is less favorable than that of the Zherebetskyy mechanism, -18.6 ±1 kcal/mol, but this needs to be considered in the context of the entire chemical reaction process. Once the necessary reaction pathways are taken in account, the alternative mechanism we describe in this section kinetically dominates the Zherebetskyy mechanism. Neither reaction has a small enough ΔH_{rxn} to be trivially reversed, and the ΔE_a of the alternative mechanism proposed here, 65 ±1 kcal/mol, is only 20% of the equivalent for the Zherebetskyy mechanism, 300

±1 kcal/mol.

3.2.6 Growth reaction mechanism

Due to the great excess of Pb to TMS_2S in the system, and the strong binding of the TMS_2S to the lead oleate hydrate described above, the initiation reaction described above would be expected to hold and then consume all of the TMS_2S in the system. This implies that the reaction most likely does not proceed at the surface of the dot as described in Zherebetskyy *et al.* Instead, it must proceed using the small PbS complexes already formed (which could be considered tiny PbS dots), since they are the only remaining source of sulfur in the solution. Therefore, the growth of the dots will act more as a solution-based polymerization of PbS quantum dot monomers than the surface reaction mechanism suggested by Zherebetskyy *et al.*

3.3 Conclusions

The Zherebetskyy mechanism for the nucleation of quantum dots is highly significant in that it was the first to take into account the effect of water in the PbS reaction, which had previously been asserted to be anhydrous. However, this reaction mechanism is incomplete in two significant ways. First, it assumes that lead carboxylate hydrate in nonpolar solution will exist as a monomer. Second, it involves a high-energy intermediate in the exchange of sulfur for oxygen in $(TMS)_2S$ becoming $(TMS)_2O$, with an activation energy of 300 ±1 kcal/mol. In this chapter, we showed that the dominant form of lead car-

boxylate hydrate in nonpolar solvent is a hydrogen-bonded dimer, and not the monomer as assumed by Zherebetskyy *et al.* We showed that the tetramer form is not favored over the dimer due to its small binding energy of 3.6 ±1 kcal/mol/monomer vs 8.5 ±1 kcal/mol/monomer for the dimer, most likely because the tetramer does not produce any new hydrogen bonds. Addressing the second issue, we provide an alternative pathway in which $(TMS)_2O$ forms the carboxylate 2(TMS)COO, with an energy barrier of only 65 kcal/mol. The small energy barrier suggests that this reaction will be limited by diffusion rather than reaction kinetics. As a result of this ease of reaction, and the fact that $Pb(oleate)_2$ is added in large excess, the amount of unreacted TMS_2S will rapidly drop to zero.

This rapid depletion of TMS_2S implies that the growth mechanism proposed by Zherebetskyy, based on TMS_2S reacting with the surface of a quantum dot, will not be a significant contributor to dot growth. By the time a quantum dot has grown to a substantial size, TMS_2S concentrations will be low. Instead, the quantum dot growth must proceed *via* smaller dots joining with larger ones in an Ostwald ripening-like process. These small nuclei, rather than the TMS_2S, should be considered the "monomers" of the reaction. The need for small dots to participate can be expected to slow and stabilize dot growth, since the passivating surface groups of the reacting dots must move out of the way before the monomers can bond. Thus, our mechanism helps to explain why dots from the Hines synthesis achieve their impressive level of monodispersity.

This new overall reaction mechanism provides information that will help us achieve thermodynamic control over the reactions that lead to PbS, and other chalcogenide, quantum dot synthesis, which will aid scale-up to an industrial

scale in which the fidelity of the structure is maintained with molecular-scale precision.

CHAPTER 4

A SIMPLE MOLECULAR REACTIVE FORCE FIELD FOR METALORGANIC SYNTHESIS

In the Hines synthesis, PbS quantum dots grow and assume a particular size distribution depending on the free energy barriers within the system. However, past the nucleation and early growth stages, these systems are too large to be studied effectively using DFT due to computational resource constraints. Chapter 2, Methods, provides a summary of DFT's limitations and the classical methods which form the alternative.

In this chapter, we present a new reactive potential to characterize the binding between a PbS nanocrystal core and lead-oxygen complexes on its surface (Figure 4). The facility with which bonds form and break between complexes and the quantum dot's surface plays a strong role in dictating the rate at which dots of varying sizes interact and, in turn, affects the resulting dot size distribution. As a result of this study, we have developed an effective reactive potential for metal-organic PbS quantum dots that we believe is general enough to be readily applied to other similar metal-organic systems, *e.g.*, other chalcogenides. Because this system performs reactions in a limited fashion between molecular units, and not for arbitrary arrangements of atoms, we call it the Simple Molecular Reactive Force Field (SMRFF). We will show that our parameterization of the SMRFF model confirms the pronounced affinity of the Pb base atoms for the OH- group, shown by Zherebetskyy *et al.* [4], and yields energy calculations in strong agreement with those provided by *ab initio* structure optimization. We also show the ability of the SMRFF approach to capture a reactive dot-dot system by direct comparison to *ab initio* results for the same reaction.

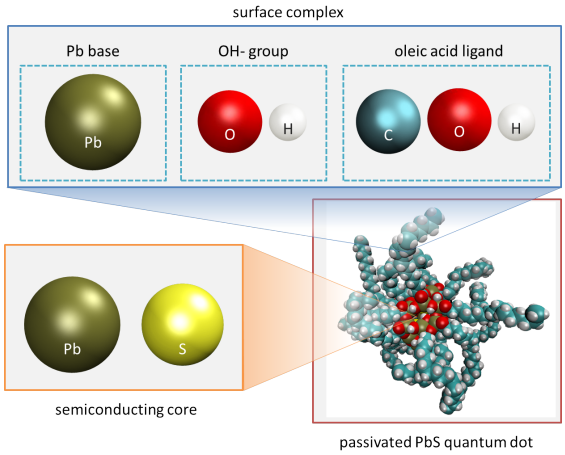


Figure 4.1: Key to the composition of our passivated PbS quantum dot systems delineated both by element (Pb, O, *etc.*) and in terms of the semiconducting PbS core and the surface complex.

The impact of this new SMRFF model is that it is perfectly positioned in terms of speed and accuracy to allow the study of the nucleation and growth of clusters and nuclei of non-traditional materials for which other force fields are not well suited and/or for which potential parameters are not available. Our immediate intent, in a subsequent study, will be to employ the SMRFF potential, developed and tested here, to study the challenge outlined above; namely, to identify the composition of the early-stage chalcogenide embryos and watch the transition from a cluster to a crystalline quantum dot.

4.0.1 Intermolecular force fields for reactive systems

Our colloidal quantum dot precursor system, a simple binary semiconductor with ligands, involves five elements (Pb, S, O, C, and H), putting it effectively beyond the reach of current parameterizations of even the best existing reactive force fields. In this work, therefore, we take a very different approach. Instead of focusing on producing a force field for a class of materials, such as metals

or hydrocarbons, we focused on the following design specifications: The force field has to incorporate the minimal set of crucial interatomic interactions that accurately portrays the system under study, and it has to be as simple as possible (with concomitantly few potential parameters) to facilitate model creation and aid computational efficiency. Our goal was to create a model that was far simpler than those previously considered to model quantum dot systems. These concepts will be described below.

Arguably, the simplest working model of a covalent bond is the Morse potential [84]. It explicitly includes the possibility of bond-breaking, and it captures the anharmonicity of the bonding. In a generalized reactive model, Morse potentials cannot be used because - as originally formulated - they do not react to bond order, and thus lead to "overcoordination." A nitrogen atom, N, for example, forms a stable triple bond with one other nitrogen atom, forming N_2. If a simple Morse model were used for the nitrogen bond, this triple bond would be unphysically repeated, producing the over-coordinated species N_3, N_4, and so on. However, in a certain class of cases, overcoordination can be prevented, even with a pairwise model: If the system does not depend on homogeneous reactions (such as N+N), and if the reactions of interest can occur only at certain locations rather than at all atoms, overcoordination can be disallowed geometrically using pairwise repulsions. For example, as a PbS crystal core forms in a quantum dot, overcoordination of the Pb atom can be prevented by increasing the S-S repulsion. Such models are not as general as Tersoff [49], COMB [54], or REAX [53], but, within their limited scope, they can be sufficiently accurate to mirror *ab initio* results, as we shall demonstrate in this chapter.

4.1 Methods

4.1.1 Functional Form

Each colloidal PbS quantum dot we studied is composed of five elements: Pb and S that constitute the semiconducting core, as well as Pb, C, O, and H, found on the surface of the dot as ligands. We define the size of each dot in terms of the number of S atoms present in the core, since the sulfur atom count is the most consistent measure of increasing size (see Figure 4.2). The Pb atom count is the other obvious measure of a quantum dot's size, but since Pb occurs both in the core and in the separable surface complexes, its quantity varies too much to be useful. Each complex has an organic component of C,O and H atoms for which developed force fields already exist, such as Jorgensen's Optimized Potentials for Liquid Simulations (OPLS) [85]. However, the suitability of OPLS is uncertain in locations directly adjacent to the Pb atom, which may strongly influence the electronic state of the neighboring atoms. Therefore, we modify the OPLS representation of the atoms nearest the Pb head of the complex, using Morse potentials for them as well as for all Pb and S atoms.

For a system of N atoms, our proposed SMRFF potential is comprised of contributions from the OPLS (E_{OPLS}) and Morse (E_{Morse}) formulations for potential energy:

$$E_{SMRFF} = E_{OPLS}\left(\mathbf{s}_1, \mathbf{s}_2 ... \mathbf{s}_N\right) + E_{Morse}\left(\mathbf{s}_1, \mathbf{s}_2 ... \mathbf{s}_N\right) \tag{4.1}$$

where $\mathbf{s}_1, \mathbf{s}_2 ... \mathbf{s}_N$ indicate the positions of the atoms in the system. Interactions

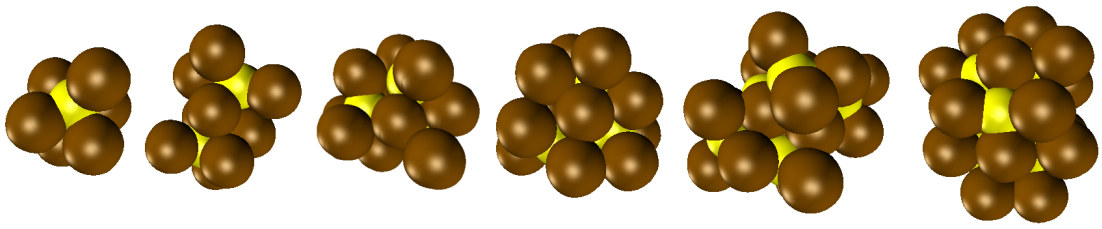


Figure 4.2: Left to right, PbS quantum dots of sizes 1-6 (larger sizes become impractical in DFT). The size is defined by the number of S atoms (yellow), since the number of Pb atoms (brown) varies with the surface structure. The organic ligands on the surface have been omitted from the image, but were included in the optimizations to determine these structures.

parameterized as part of our custom potential substitute the Morse potential in place of Lennard-Jones (thereby setting the Lennard-Jones parameters to zero), and are primarily those which involve inorganic atom types. Table 4.2 details the potential function used for each interaction.

In what follows below, $\mathbf{s}_i = x_i\hat{\mathbf{i}} + y_i\hat{\mathbf{j}} + z_i\hat{\mathbf{k}}$ is the position vector of atom i in a system of N atoms.

The OPLS description of the potential energy takes the form:

$$E_{OPLS}\left(\mathbf{s}_1, \mathbf{s}_2 ... \mathbf{s}_N\right) = E_{bonds} + E_{angles} + E_{dihedrals} + E_{nonbonded} \tag{4.2}$$

$$E_{bonds} = \sum K_r \left(r - r_0\right)^2 \tag{4.3}$$

$$E_{angles} = \sum K_\theta \left(\theta - \theta_0\right)^2 \tag{4.4}$$

$$
\begin{aligned}
E_{dihedrals} = \quad & \tfrac{V_1}{2}[1 + \cos(\phi - \phi_0)] \quad + \\
& \tfrac{V_2}{2}[1 - \cos(2(\phi - \phi_0))] \quad + \\
& \tfrac{V_3}{2}[1 + \cos(3(\phi - \phi_0))] \quad + \\
& \tfrac{V_4}{2}[1 - \cos(4(\phi - \phi_0))]
\end{aligned}
\tag{4.5}
$$

$$
E_{nonbonded} = \sum_{i>j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{k_e q_i q_j}{r_{ij}} \right) \tag{4.6}
$$

$$
A_{ij} = \sqrt{A_{ii} A_{jj}} \quad \& \quad B_{ij} = \sqrt{B_{ii} B_{jj}} \tag{4.7}
$$

In the above equations, E is the potential energy, E_{bonds} is the bond-stretching energy, E_{angles} is the angular-dependent energy, $E_{dihedrals}$ is the dihedral twisting energy, $E_{nonbonded}$ contains the pairwise van der Waals and Coulombic energy. In equation (4.3), K_r is the spring constant of a given bond (assumed to be harmonic), the bond length r is a function of position vector pairs $\mathbf{s}_i, \mathbf{s}_j$, with equilibrium value r_0. In equation (4.4), k_θ is the spring constant of a given angle (assumed to be harmonic), the angle θ is a function of position vector triplets $\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k$, with equilibrium value θ_0. In equation (4.5), $V_1 - V_4$ are the energy coefficients of the cosines of increasing frequency which define the dihedral interaction, ϕ is the torsional angle about the dihedral with an eclipsed configuration = 0, and is a function of four atom position vectors $\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_l$, with equilibrium value ϕ_0. In equation (4.6), A_{ij} and B_{ij} are the Lennard-Jones coefficients of the i-j pair interaction, r_{ij} is the pair distance, k_e is the Coulomb constant, q_i and q_j are the atomic charges. Finally, in equation (4.7), A_{ii}, A_{jj}, B_{ii}, and B_{jj} are the Lennard-Jones parameters of pure i-i and j-j interactions.

The Morse potential for an atom pair is defined as

$$E_{Morse} = D_0[e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}] \tag{4.8}$$

where D_0 is the equilibrium potential energy, r_0 the equilibrium separation distance, and α the stiffness of the resulting bond. It is a function only of the separation distance of an atom pair with positions $\mathbf{s}_i, \mathbf{s}_j$.

While OPLS parameters are available for the charges on hydrogen, carbon and oxygen atoms, the charges of atoms close to the nanocrystal surface are not part of standard force fields, and we parameterized them alongside the Morse potential parameter set. Although it would also be possible to set such charges *via ab initio* population analysis, charge density calculations are not designed to minimize the error of the resulting potential energies across the parameter set, unlike our optimized charges.

4.1.2 SMRFF Parameterization Method

In addition to choosing a suitable functional form, as described above, assembling a sufficiently accurate and extensive reference data set is critical in determining the force field parameters. A data set that encompasses a broad range of possible conditions and configurations of the system is likely to improve the parameterized force field. However, it is also important to strike a balance between the breadth of data that can be reproduced and the computational time that is necessary to generate that data set. Given these constraints, we identified several key system properties as targets to reproduce, which guided the process of data collection and assembly.

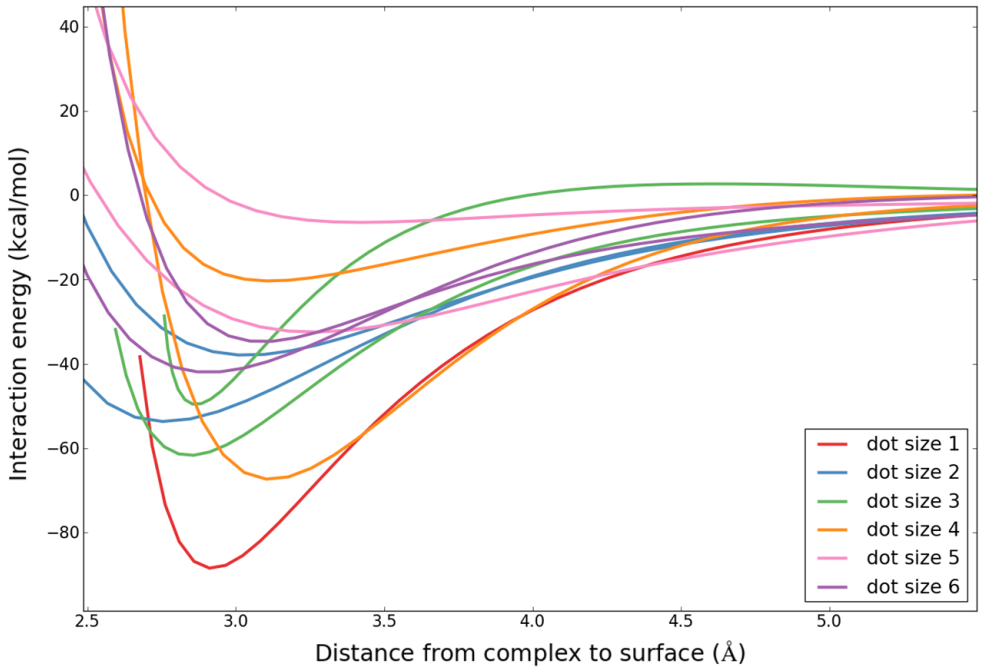


Figure 4.3: Energy required to remove a given surface complex from the rest of the dot, for dot sizes 1-6, identified by the color key in the inset. Note that there is no simple relationship between dot size and binding energy, and that dots can have multiple distinct binding energies corresponding to removal of each of their possible surface complexes. In order to be useful, any model must reproduce this extraordinary variability.

First, we targeted the interaction energies between the different components of the system: the surface complexes and the quantum dot core (Figure 4.3). For the complex-to-core and complex-to-complex interactions, our example systems consisted of *ab initio* trajectories of the separable complexes (ligands) moving with respect to the surface and each other. For the core-to-core interactions, our example systems involved trajectories of parts of the core moving with respect to each other. These interaction energies are significant because they create the energy barriers that determine the properties of the system.

Second, we targeted the forces on all of our hundreds of DFT-optimized quantum dot systems. At these states, the RMS force on the atoms is close to zero (within the convergence criterion for the optimization, 3.0×10^{-4} Hartrees/Bohr). Our parameterization sought to also produce net forces of zero at these geometries, thereby increasing the accuracy of our model in the vicinity of the stable and metastable states which will be sampled most heavily during simulations. We found that it was especially necessary to include optimized structures which featured an excess of complexes near the surface, in order to encourage the parameters to reject overcoordination of the lead atoms.

All of the structural simulations in the data set were performed using the quantum chemical software package Gaussian 09 [63]. Energy values were determined *via* single-point energy calculations using the HSE06 functional [65, 67, 66] and LANL2DZ basis set+core potentials [69]. The HSE06 hybrid functional was selected because it is a member of the PBE family of hybrid functionals, widely felt to be effective for describing metal complexes [66, 43], and because it is also range-separated, making it much more accurate for systems like ours which feature long-range Coulombic interactions [65]. Our calculations show that the HSE06/LANL2DZ combination is the most effective of the 24 functionals that we tested at predicting the Pb-Pb dimerization energy while still decaying slowly at long range. We included the effect of the toluene solvent in the DFT calculations *via* using Gaussian 09's default Self-Consistent Reaction Field model, the Polarizable Continuum Model [73, 63], which creates a charged cavity around the atoms that affects the electron density via its electrostatic interactions.

In order to select our DFT model for the system, we considered a number of

available levels of theory in DFT, in terms of both the most appropriate functional and basis set. The functionals we considered were M06L, PBE0, HSE06 and HISS [86, 87, 65, 88], known in Gaussian as M06L, PBE1PBE, HSEH1PBE and HISSbPBE, respectively. We also included MP2 and CCSD, [89, 90] although these calculations were not successful. The basis sets we considered were CEP-31G, CEP-121G, LANL2DZ, and Def2TZVP. These basis sets also include pseudopotentials on the heavy atoms, particularly lead. We compared the binding energy of the neutral Pb-Pb dimer in vacuum (see Figure 4.4) using a BSSE correction. As a result of these calculations, we selected the HSE06 functional in combination with the LANL2DZ basis set for the accurate binding energy that this produced, and its slow decay with distance, counteracting the known tendency of DFT to decay too rapidly. [41]

The importance of selecting an appropriate functional is demonstrated by Figure 4.5. The binding energy difference between two functionals, say between M06L (pure DFT) and HSE06 (a range-corrected hybrid) can be up to 8 kcal/mol for our system.

From energy calculations, the interaction energy between two rigid structures A and B may be calculated using the following equation:

$$E_{inter} = E(AB)^{AB} - E(A)^{AB} - E(B)^{AB} \tag{4.9}$$

Here, AB represents a group of atoms, A being one subset of the atoms and B being the remaining atoms. The superscripts indicate that ghost atoms are used to ensure that A, B, and AB are all evaluated using the same number of basis functions, to avoid BSSE [81] as described in the previous chapter. Figure

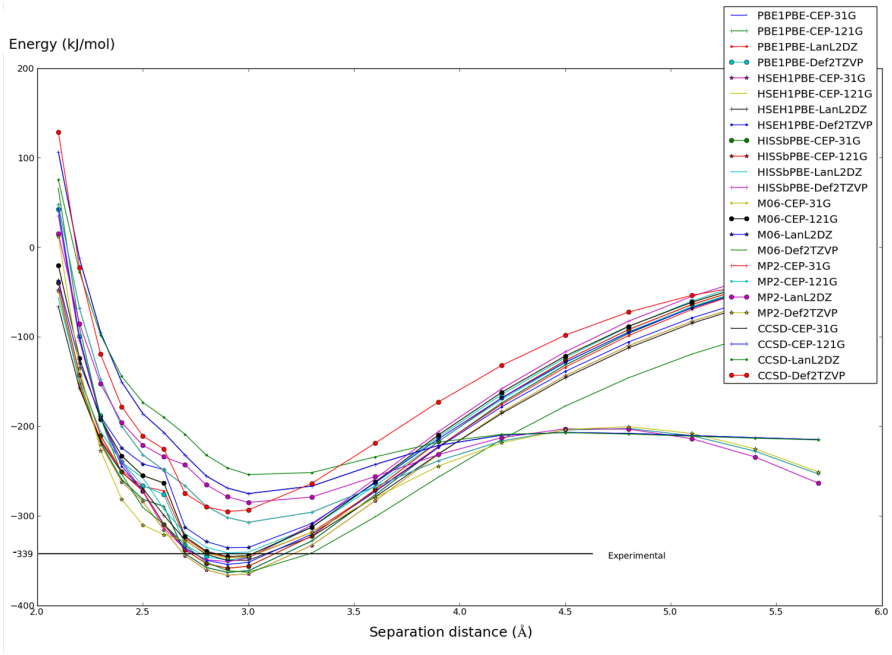


Figure 4.4: *Ab initio* Pb-Pb dimerization at different levels of theory, compared with experiment. We selected HSE06/LANL2DZ, with an accurate minimum of 345 kJ/mol and slow long-range decay, avoiding the characteristic error of DFT in which interactions decay too quickly with distance [41].

4.6 illustrates the importance of this step by showing the uncorrected and corrected energy data. The uncorrected calculations show systematically deeper energy minima because their fitting of the electron wavefunction becomes less accurate as the interatomic distances increase, thus creating spuriously high energies for the more separated atoms. The resulting 9 kcal/mol difference would be significant if not corrected.

The *ab initio* geometries and interaction energies for dot sizes 1 to 6, ranging from 52 to 198 atoms, comprised the final data set. Larger dots, of size greater than 200 atoms, would be impractical in DFT. Tables 4.1 and 4.2 summarize all the atom types and corresponding interaction types considered in the parameterization. Note that, due to their ability to participate (or be found in) differ-

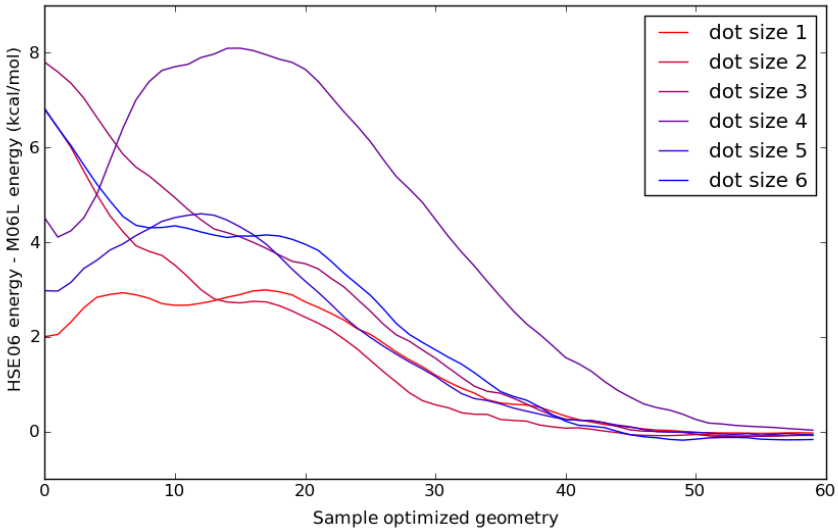


Figure 4.5: Comparison of interaction energies derived from several single point energy calculations using the M06L and HSE06 functionals. The plot shows the difference in energy between the two functionals as a function of different system geometries, each along a trajectory of removing one complex from the nanocrystal surface, color-coded for dot sizes 1-6 as defined in the inset.

ent local atomic environments, atoms of the same element in different environments had to be distinguished in the parameterization scheme. One such case concerned the Pb present in the core, denoted PbS, and that in the complexes, denoted PbO. The first element symbol indicates the element of interest, while the subsequent symbol or symbols characterize the bonding environment. This helps to distinguish among atoms of the same element type treated differently in the parameterization.

The increasing complexity of the PbS quantum dot as it grows in size places limitations on the scope of DFT calculations that can be performed. We reduced the computational cost by recognizing that the tails of the long oleic acid chains do not participate in the surface interactions, which are dominated by atoms in close proximity to the core, within the range of the interatomic forces. Accord-

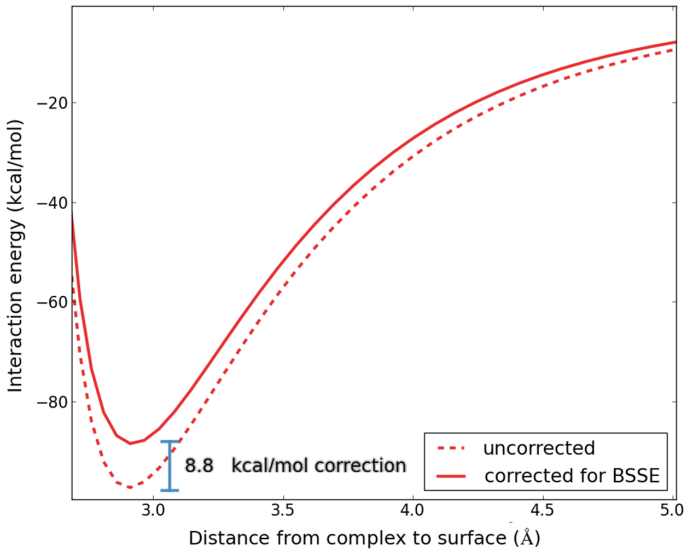


Figure 4.6: BSSE-corrected binding energy curves (solid) vs. BSSE-uncorrected curves (dashed). The uncorrected curves are always deeper because BSSE causes overestimation of binding energies.

Abbreviation	Name
PbS	Core Pb
S	Core S
PbO	complex Pb
OH	complex oxygen in hydroxide ion
HO	complex hydrogen in hydroxide ion
OO	complex oxygen in ethanoic acid base
COO	complex carbon in ethanoic acid base
CCO	complex carbon in ethanoic acid tail
HC	complex hydrogen in ethanoic acid tail
M	Morse Potential
LJ	Lennard-Jones Potential

Table 4.1: Relevant abbreviations

ingly, we substituted the oleic acid ligands (used experimentally) with shorter ethanoic acid chains in our simulations as we formulated surface interaction parameters. Beyond the non-conductive C-C single bond, the electronic states

of the ligand are unlikely to affect those of the reactive site. [76] On the other hand, oleic acid tails contribute to the off-surface interactions and may be effectively described by the existing OPLS force field. Within the neighborhood of more unpredictable surface interactions, our parameterization introduces custom Morse potentials paired with electrostatic forces enabling us to capture surface phenomena more precisely.

4.1.3 Parameter Optimization

From the *ab initio*-optimized structures of dot sizes 1 to 6, we computed histograms of the interatomic distances, allowing us to set the initial parameter values in a systematic fashion. We speculated that a collection of distances classified by interaction type, such as Pb+S, would offer insight into the preferred separation distance of each, a technique suggested by Kirkwood for fluid systems [91] (Figure 4.7). Since we need only initial guesses, which we will then begin to optimize, we feel justified in applying the theory more generally to semi-crystalline solids.

The parameter optimization process may be divided into two steps: a global search of the working parameter domains and a local convergence towards optimal values for each parameter. These steps occur successively, but also operate cyclically (essentially a feedback loop), relaying information from the previous iteration that allows for more informed subsequent iterations. The error function to be minimized is:

$$Error = \sum^{N} \left(E_{binding,DFT} - E_{binding,SMRFF} \right)^2 + 0.3 \times \sum^{M} \left(F_{complex,SMRFF} \right)^2 \qquad (4.10)$$

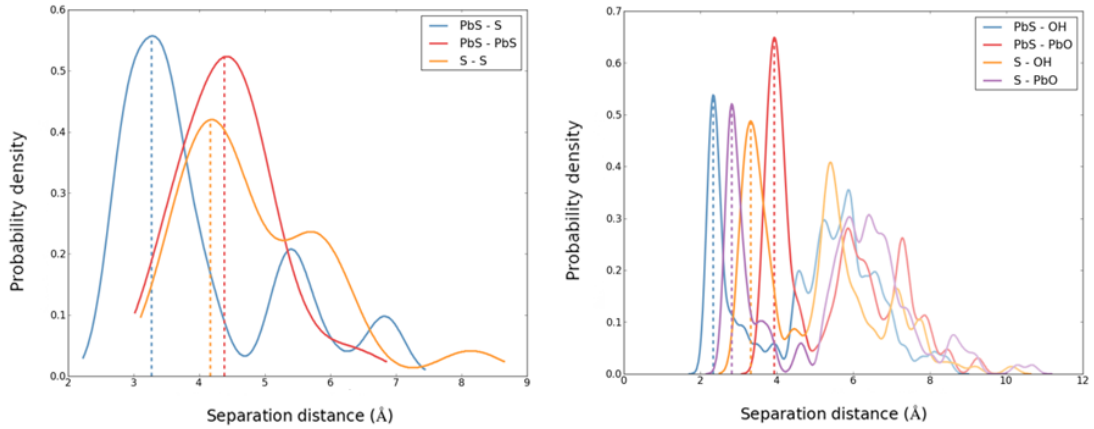


Figure 4.7: The radial distribution functions between different atom types were used to provide initial parameters for their interactions. The location of the first peak of each curve (denoted by vertical lines) is the likely equilibrium distance for each pair. Left: interactions within dot cores; right: interactions between surfaces and cores. Color code as defined in the insets.

where N is the number of complexes bound to the dot surfaces in the entire data set and M is the number of complexes bound to the dot surfaces in the geometry-optimized section of the data set. $E_{binding,DFT}$ is the interaction energy between the complex and the rest of the dot according to DFT, and $E_{binding,SMRFF}$ is the same energy according to the SMRFF model being parameterized. $F_{complex,SMRFF}$ is the magnitude of the net force on each DFT-optimized complex according to the SMRFF model. The units of energy are kcal/mol and the units of force are kcal/mol/Angstrom.

Each parameter was constrained to a certain range in order to enforce physical realism, *e.g.*, the Pb-Pb Morse radius r_0 was not permitted to fall below 3.0 Angstrom. The global search was performed by initially relaxing these constraints on each parameter, then confining them as consistencies in the behavior of certain parameters emerged. However, the variability of a large number of

parameters remained high and called for further partitioning of the problem. To address this, we focused on smaller subsets of parameters at a time. For example, we used DFT to evaluate dots with only one attached ligand on each, rather than a full coating of ligands. When implemented in the data fitting, this change had the effect of placing greater emphasis on core-complex interactions, temporarily relegating complex-complex interactions into the background. Once the parameters describing core-complex interactions appeared to converge on accurate values, the parameter set was transferred to the data of complete quantum dot structures for an adjustment of complex-complex interaction parameters. The core-complex parameters were still allowed flexibility during this step, but conformed to stricter bounds. Details of these bounds and the completion criteria are explained in detail below.

With regards to the local optimization, a number of options are available for performing a non-linear curve fit, commonly divided into gradient-based or derivative-free algorithms. We investigated several implementations and selected the local derivative-free algorithm "subplex," an adaptation of the Nelder-Mead simplex algorithm, [92] in the NLopt nonlinear optimization library [93]. The optimization minimizes the sum of squares of residuals between the data and the fitted function evaluation, and also has the advantage of increasing computations linearly with problem size, favorable for a problem of our scope. [92] We monitored the parameters as they changed to prevent any issues with their physicality. Thus, our approach combined both algorithmic optimization and active oversight of the parameterization.

For fully optimized structures, the vector sum of forces on each complex must sum to zero, reflecting a state approaching equilibrium. This fact imposes

an additional target for the parameterization. In order to implement this concept, the corresponding forces for each proposed potential function were calculated alongside energy evaluations during the parameterization process. Using the fact that the force F is defined as the negative gradient of the potential energy, we obtain the following equations for the forces:

$$F_{Morse} = 2\alpha D_0[e^{-2\alpha(r-r_0)} - e^{-\alpha(r-r_0)}] \tag{4.11}$$

$$F_{LJ} = \frac{24\epsilon}{r}[2(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6] \tag{4.12}$$

$$F_{coul} = kq_1q_2/r^2 \tag{4.13}$$

where F_{Morse}, F_{LJ}, and F_{coul} are the Morse, Lennard-Jones, and Coulombic forces, r is the interatomic distance, $alpha$, D_0, and r_0 are the Morse parameters, ϵ and σ are the Lennard-Jones parameters, and k is the Coulomb constant, and q_1 and q_2 are the atomic charges.

One aspect of the minimization which was at our discretion was the relative weight attributed to the energy calculations and force calculations. Both were incorporated in determining the error and hence determining the optimal parameterization to minimize that error. Preliminary error goals were established to guide this process. For the forces, we consulted the convergence criteria of the Gaussian 09 software with which the quantum dot structures were optimized, which requires a root mean squared (RMS) force per atom below 3.0×10^{-4} Hartrees/Bohr for a normal optimization and an RMS force below 1.7×10^{-3} Hartrees/Bohr for loose optimization. For the 10-atom complex, this yields an RMS error target below 20 kcal/mol/Angstrom. A scaling factor of 0.3 applied to the force error caused the optimization to satisfy this force con-

straint. We found that a scaling factor of 0.5 or 1.0 caused the parameterization to neglect the energy error, while lower scaling factors caused the optimization not to meet the desired force constraint.

	PbS	S	PbO	OH	OO	HO	COO	CCO	HC
PbS	M	M	M	M	M	M	M	M	LJ
S	M	M	M	M	M	M	M	M	LJ
PbO	M	M	M	M	M	M	M	M	LJ
OH	M	M	M	M	M	M	M	M	LJ
OO	M	M	M	M	M	M	M	M	LJ
HO	M	M	M	M	M	LJ	M	LJ	LJ
COO	M	M	M	M	M	M	M	M	LJ
CCO	M	M	M	M	M	LJ	M	LJ	LJ
HC	LJ	LJ	LJ	LJ	LJ	LJ	LJ	LJ	LJ

Table 4.2: A compilation of all interaction types and the potential function parameterized to handle each interaction. Note that electrostatic interactions were included for all interactions in addition to the appropriate potential shown in this Table. A key of all abbreviations is provided in Table 4.1.

Tables 4.3 through 4.5 detail all the interatomic potentials used in the final parameterization scheme. The Lennard-Jones parameters are referenced from existing OPLS parameters [85], applying the following geometric mixing rules:

$$\epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \tag{4.14}$$

$$\sigma_{AB} = \sqrt{\sigma_A \sigma_B} \tag{4.15}$$

where ϵ_{AB} is the Lennard-Jones energy parameter for the pair AB, ϵ_A and ϵ_B are the energy parameters for the homogeneous pairs AA and BB, and σ_{AB}, σ_A, and σ_B are likewise the Lennard-Jones distance parameters for the mixed pair AB and the homogeneous pairs AA and BB.

In cases here Lennard-Jones parameters were zero or unavailable, these atom pairs were given weak interactions (ϵ=0.01 kcal/mol, σ=2.5 Angstrom) for greater stability in MD simulations. Most of these cases occur for interactions with hydrogen. A cut-off of 10 Angstrom was imposed on Morse and Lennard-Jones interactions due to their rapid decay with increasing separation distance [47]. The long-range electrostatic potential was calculated pairwise for all charges, with no cut-off.

4.1.4 Evaluation of Error

The simplest measure of the performance of our model is the root-mean-square (RMS) error of the energies and forces of our model with respect to the *ab initio* data set. Given that the data set is sufficiently comprehensive, the RMS error can be seen as the expected standard deviation of the errors of the model in actual use. In addition to assessing RMS error, we also analyzed the structural fidelity of PbS quantum dots simulated using our force field as compared to those optimized *ab initio*. For this purpose, we generated a set of optimized quantum dot structures using our model, re-optimized these states using DFT, then quantified the structural shift (which ideally would be zero) between the optimum states of the two models. We also compared the structural consistency more globally by revisiting a technique employed previously, namely, analyzing histograms of the interatomic distances. This takes into account the lattice constants and all the bond lengths of the key interactions. To develop a well-populated sample space for the probability density distribution, we generated a collection of optimized quantum dots using our parameterized force field. Our comparison set was composed of the DFT-optimized structures used in the pa-

Table 4.3: Fitted Morse Potential Parameters

Interaction	D (kcal/mol)	$\alpha(^{-1})$	$r_0()$
(PbS,PbS)	3.147	0.422	7.500
(S,S)	0.875	0.629	6.600
(PbS,S)	33.094	2.086	3.204
(PbS,OH)	11.200	1.850	2.331
(PbS,PbO)	0.010	1.011	7.091
(S,OH)	0.001	1.101	6.542
(S,PbO)	9.636	1.739	2.891
(PbS,OO)	1.465	1.251	2.716
(S,OO)	0.012	0.747	7.500
(PbO,OH)	25.086	2.308	2.130
(PbO,PbO)	0.007	0.938	7.500
(OH,OH)	0.001	2.400	4.045
(PbO,OO)	8.694	2.316	2.134
(OO,OH)	0.001	2.298	4.365
(OO,OO)	0.500	2.400	2.000
(OO,COO)	0.819	1.810	3.635
(OO,CCO)	0.001	1.438	5.800
(OH,COO)	0.100	2.500	1.947
(OH,CCO)	0.001	2.400	2.200
(PbS,COO)	0.500	1.975	3.549
(PbO,COO)	0.010	2.388	4.389
(S,COO)	3.850	2.150	3.909
(PbS,CCO)	3.101	2.200	3.328
(PbO,CCO)	3.635	2.100	3.372
(S,CCO)	0.225	2.409	3.600
(COO,COO)	0.500	2.151	3.260
(PbS,HO)	0.076	1.225	4.500
(PbO,HO)	0.007	1.570	4.900
(S,HO)	0.010	2.200	3.675
(OH,HO)	2.368	1.625	1.800
(OO,HO)	0.379	1.701	2.200
(COO,HO)	0.081	2.401	3.589
(COO,CCO)	3.804	2.500	2.745

rameterization process. For each data set, we calculated all observed separation distances for a particular interaction type, such as Pb+S, and fit them to a probability density distribution. The first peaks of the curves were recorded, as they denote the most probable equilibrium separation of each pairwise interaction.

Table 4.4: Lennard-Jones Potential Parameters

Interaction	ϵ (kcal/mol)	σ()
(CCO,CCO)	0.066	3.500
(CCO,HO)	0.000	0.000
(HO,HO)	0.000	0.000
(HC,HO)	0.010	2.500
(HC,OH)	0.010	2.500
(HC,OO)	0.010	2.500
(HC,COO)	0.056	3.06
(HC,CCO)	0.044	2.96
(HC,HC)	0.030	2.500
(HC,PbO)	0.010	2.500
(HC,PbS)	0.010	2.500
(HC,S)	0.010	2.500

Table 4.5: Charges

Atom	Charge (e)	Atom	Charge (e)
PbS	1.124	OO	-0.698
S	-1.124	HO	0.370
PbO	0.865	CCO	-0.280
OH	-0.726	HC	0.060

This comparison is shown in Figure 4.11 in Section 4.2.

4.1.5 Implementation of a Reactive Dynamics Demonstration

We present a preliminary implementation of the SMRFF model to recover the potentials of mean force (PMFs) for a sample of PbS quantum dot interactions. We have used the collective variables module [94] available in the LAMMPS Molecular Dynamics Simulator [95] to simulate the reaction of two PbS quantum dots whose interaction is governed by the SMRFF model. For this reaction, there are many possible start and end points, and so a chain-of-states method such as Nudged Elastic Band [77] is less viable than a reaction coor-

dinate method such as metadynamics [94]. The selection of an appropriate reaction coordinate is nontrivial: we selected a distance-based variable as our reaction coordinate due to the strong dependence of interaction energy on the proximity of the two interacting dots. The particular distance-based variable we chose was the radius of gyration, R_{gyr}, defined as the root mean square of the distances of the N atoms from the system's center of geometry:

$$R_{gyr} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\vec{x_i}(t) - \vec{x_{cog}}(t)\|^2} \tag{4.16}$$

where $\vec{x_i}(t)$ is the position of atom i at time t and $\vec{x_{cog}}(t)$ is the position of the system's center of geometry at time t. Unlike other distance-based variables such as the distance between the centers of mass of each quantum dot, the radius of gyration does not privilege any particular division of the system into parts. This makes it appropriate for sampling the final stages of the reaction, in which all the atoms of both original dots should be allowed to interact freely.

Different system configurations, or "frames," of a single metadynamics trajectory were sampled to conduct a validation of the energies using DFT. For each of the sampled frames, a geometry optimization was performed in Gaussian 09 [63] holding the separation of the S atoms fixed. Note that only one S atom is present in each dot for an interacting system of two dots of size 1; thus only one constraint is imposed. We performed the calculations using both the HSE06 functional and the M06L functional for comparison, with the LANL2DZ basis set for both. We included the effect of SMD toluene solvent as in our previous DFT calculations. Of the ten metadynamics simulations performed using the collective variables module, the two whose trajectories achieved the lowest global potential energy were used in the validation.

The Python and C++ code used for all the above calculations can be accessed on Github at https://github.com/jminuse/simple-molecular.

4.2 Results

4.2.1 SMRFF Validation

Our model was able to reproduce, largely within tolerances of ±2 kcal/mol, the many distinct binding energy curves of the complex-to-dot system (Figure 4.8). Interestingly, the largest deivation of 10 kcal/mol is for the smallest dot, size 1, rather than there being a trend for the error to worsen with larger dot sizes. This is helpful because we plan to extrapolate toward larger sizes with the force field, so it is useful that the error remains in check with increasing dot size.

It is interesting to compare the binding energies of the complexes to the dots, at most -80 kcal/mol, with the much greater binding energy we calculated for the ligands to the complex, -360 kcal/mol. This strongly suggests that it is the whole complex which is the unit that can leave the dot, as in our model - the organic ligands are not capable of departing on their own and leaving the surface unpassivated. This affinity of the Pb atoms for the OH- group confirms the results of Zherebetskyy *et al.* [4].

The core-to-core interaction energies are shown in Figure 4.9. These results show the SMRFF energies remaining within the ±2 kcal/mol band reliably, whether for small or large dots. Once again there is no trend toward greater deviations at larger dot sizes. The interaction energies grow with increasing dot

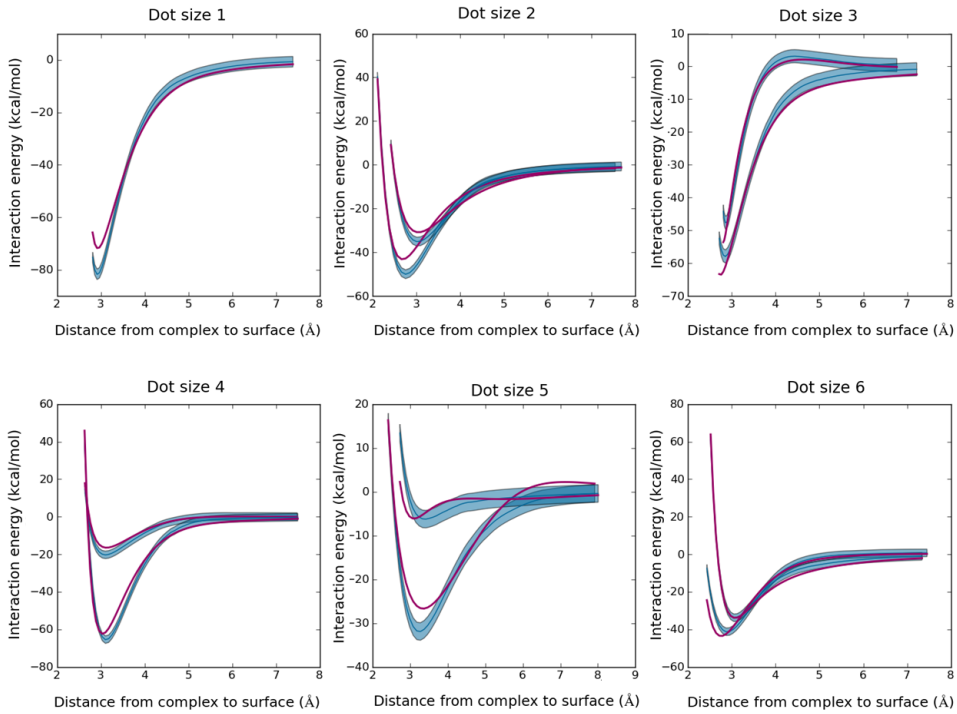


Figure 4.8: Complex-PbS surface interaction energies by dot size (from 1-6) as calculated using Gaussian 09 (shown as a blue band ±2 kcal/mol wide, see Supporting Information) and with the corresponding SMRFF predictions in red. Each curve corresponds to one complex (a ligand) being pulled away from the corresponding dot.

size because the larger dots have more core atoms to interact with each other.

The energies for overcoordinated systems are shown in Figure 4.10. Overcoordinated systems are particularly difficult to model accurately with pairwise potentials (such as the Coulombic, Lennard Jones, and Morse potentials use in SMRFF) because pairwise potentials have no explicit environment-dependent terms to indicate that an atom is sufficiently coordinated or overcoordinated. Since this is a very typical failure mode for pairwise models, a pairwise model cannot be considered valid until overcoordinated systems are taken into ac-

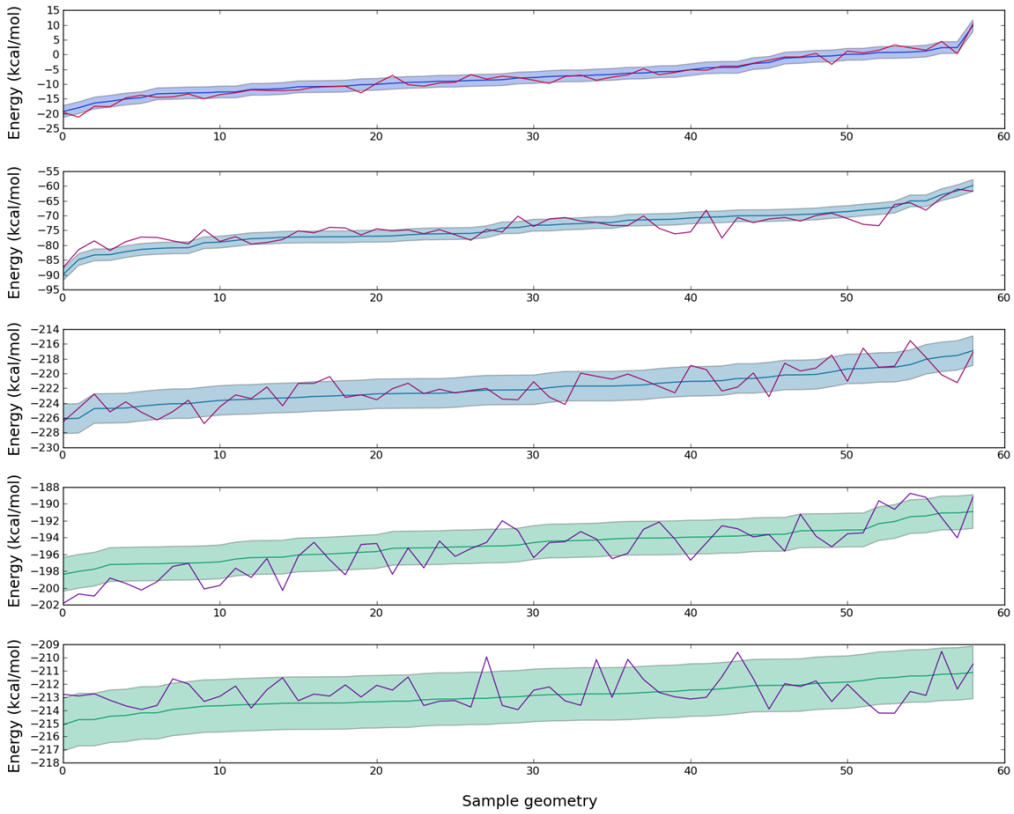


Figure 4.9: PbS-PbS core interactions calculated using Gaussian 09 (shown as a green band ±2 kcal/mol wide) and the corresponding SMRFF predictions in red. We define the core interaction as the binding energy between two halves of the core if it were divided into two charge-neutral pieces. Since larger cores can have multiple ways of dividing the core in half, we show one division of dot size 2 (top graph) and two divisions each of dot size 3 and 4 (second pair and bottom pair, respectively).

count. Using O-O, S-S, and Pb-Pb repulsive terms to prevent these atoms from clustering excessively, we correctly modeled overcoordination in PbS quantum dots.

The root mean squared errors in energy resulting at the end of the parameterization are presented in Table 4.6 and Table 4.7. Root mean squared errors for force are shown in 4.8.

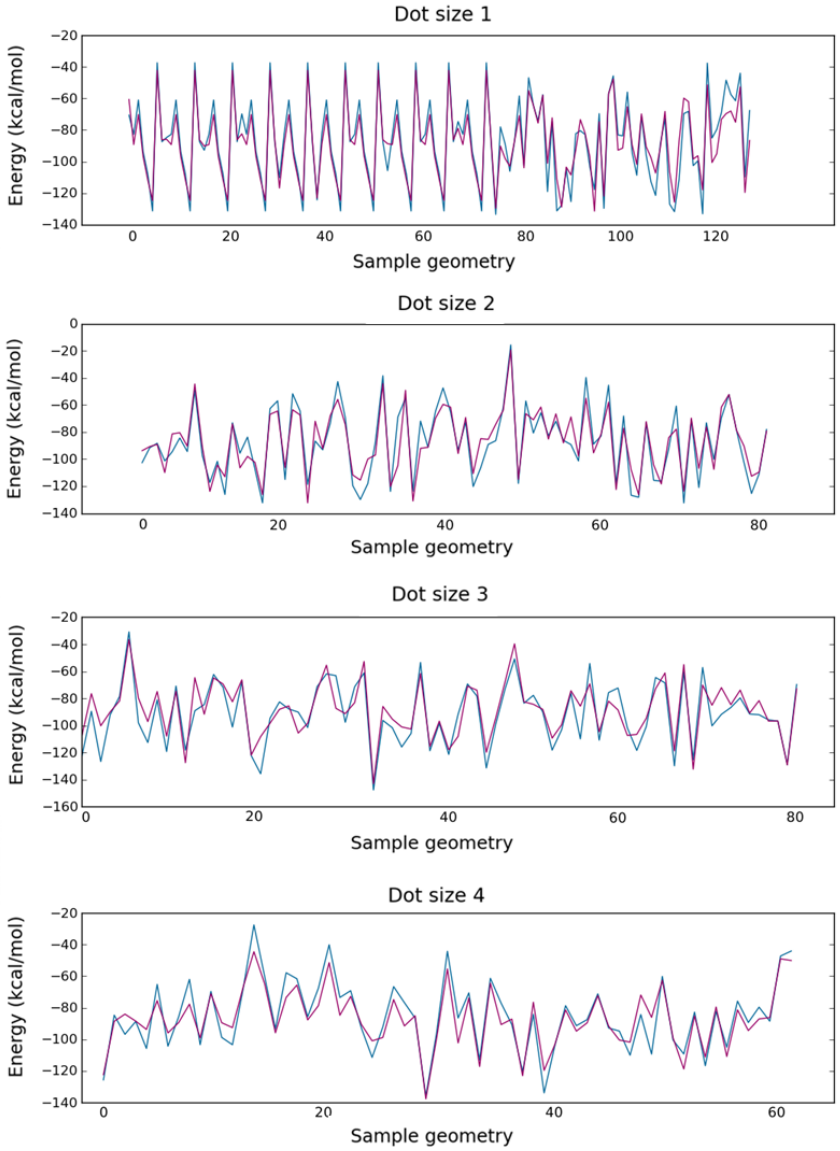


Figure 4.10: Gaussian 09-generated data (shown in blue) for several complex-to-nanocrystal interaction energies derived from overcoordinated structures, with the corresponding SMRFF fit in red, as a function of 359 distinct single-point calculations derived from several overcoordinated geometries. Results are shown for dot sizes from 1 (top plot) to 4 (bottom plot) because overcoordinated systems require more atoms than simply passivated systems, making overcoordinated dots of size 5 and 6 impractical in DFT.

Table 4.6: Root mean squared errors in energy for complex-to-dot and core-to-core interactions

Complex-to-dot interactions	RMS error ($kcal/mol$)
composite	2.8
dot size 1	3.7
dot size 2	2.8
dot size 3	2.3
dot size 4	2.6
dot size 5	2.7
dot size 6	3.0

Core-to-core interactions	RMS error ($kcal/mol$)
composite	1.8
dot size 2	1.3
dot size 3	2.1
dot size 4	1.6

The model was also able to reproduce the structural properties of the dots accurately: Structural comparisons between dots optimized with our model and with DFT predictions are given in Figure 4.11 and in Table 4.9.

Figure 4.11 has to be interpreted with care because it shows pairwise distances, which are relatively easy for a pairwise potential to match accurately. This type of comparison can be used spuriously to cover up poorly-parameterized pairwise models (for example, given the PbS cubic lattice constant of 0.6 nm, a disastrous pair potential which reduced the crystal to a PbS nano-chain would still appear correct as long as the Pb-Pb chain length were 0.6 nm). The results shown in Table 4.9 are more impressive because they show mean absolute

Table 4.7: Root mean squared errors in energy for overcoordinated systems

Overcoordinated structures	RMS error (*kcal*/*mol*)
interaction energy > −108.6 kcal/mol	(75% of data)
composite	8.5
dot size 1	7.3
dot size 2	9.5
dot size 3	9.5
dot size 4	7.8
interaction energy > −147.7 kcal/mol	(all data)
composite	9.2
dot size 1	8.2
dot size 2	9.9
dot size 3	10.7
dot size 4	8.3

Table 4.8: Root mean squared net forces on the complexes in optimized systems. Our target RMS net force, based on the ”loose” optimization convergence criterion in Gaussian 09, was 20 kcal / mol / Angstrom [63].

Force on one complex	RMS force (kcal/mol/Angstrom)
composite	19.6
dot size 1	21.4
dot size 2	23.0
dot size 3	16.8
dot size 4	20.9
dot size 5	18.6

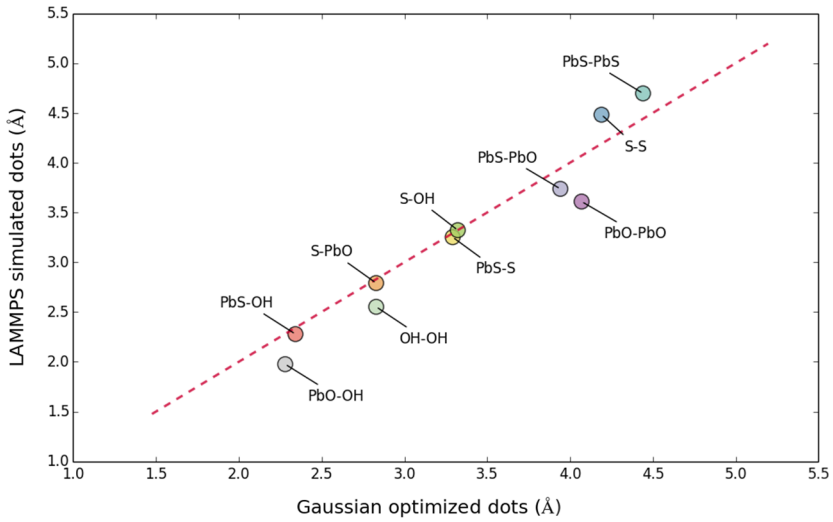


Figure 4.11: The most frequent interatomic distances for each interaction in our model, plotted against the equivalent distances in the DFT data set. The dashed line shows the ideal 1:1 ratio. The mean deviation from DFT distances is 6%.

variations between atom positions for two different structures, a non-pairwise property which cannot be as easily faked by poor structures. Orthogonal Procrustes rotation is used to remove rigid translation and rotation so that the two structures can be compared.

4.2.2 Reactive Dynamics

Figure 4.12 displays the PMFs of the two metadynamics trajectories selected for analysis. For the first reaction path, sample configurations are also shown at their corresponding radii of gyration for a visualization of the system. For comparison with the DFT calculations, it is necessary to calculate the potential energy of the system, as shown in Figure 4.13. Note that, unlike the PMFs, each potential energy curve is not constructed from aggregate information on

Table 4.9: Mean absolute variations in atomic positions between dots optimized with SMRFF and DFT.

Dot size 1	mean absolute variation (Angstrom)
core	0.48
complexes	0.36
complete structure	0.91
Dot size 4	**mean absolute variation (Angstrom)**
core	0.39
complexes	0.36
complete structure	0.93

the trajectory as a whole but, rather, is available after a pre-specified number of timesteps in the simulation. Even for small variations, $\delta\xi$, in the reaction coordinate, the potential energy exhibits a large spread, as evident from Figure 4.13; hence, we applied a smoothing filter to improve the interpretation of the average behavior at a given value of the reaction coordinate.

Sample configurations for performing the DFT calculations were selected to target a broad range of reaction coordinate values, and the S atom separation was fixed to preserve similar values of the reaction coordinate during optimization. Not all optimizations met each of the targeted Gaussian 09 criteria for convergence; hence, we provide results for the energies of structures whose root mean squared (RMS) force converged to within 1.7×10^{-3} Hartrees/Bohr, the Gaussian 09 loose optimization criteria for RMS forces [63]. 17 of 20 sampled configurations from the first trajectory, and 14 of 16 sampled configurations from the second trajectory, satisfied this condition when the HSE06 functional

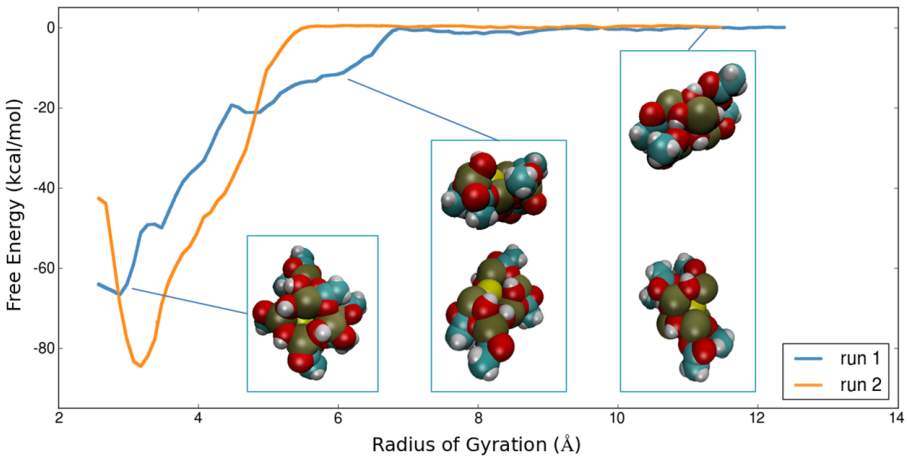


Figure 4.12: Potentials of mean force for two metadynamics trajectories, 1 (blue) and 2 (orange). The inset images show the dots at the corresponding radii of gyration in trajectory 1.

was used. For calculations using the M06L functional, 13 of 16 and 6 of 11 configurations met the RMS force convergence criteria for trajectories 1 and 2, respectively.

Figures 4.14 and 4.15 display the DFT-calculated energies together with results from Molecular Dynamics using metadynamics for trajectories 1 and 2, respectively. Each set of data is graphed such that the largest available radius of gyration for that dataset coincides with zero energy. Results using both the HSE06 and M06L functionals are provided for comparison. The large difference between the results from different functionals reflects the fact that even Density Functional Theory is sensitive to the choice of model. The M06L potential energy curve shows a consistent bias toward underpredicting the reaction energy, as expected since the HSE06 functional is range-separated and thus, for spatially extended and nonbonded systems, superior to a pure-DFT functional such as M06L in the absence of dipole-dipole or dispersion correction [65]. When the

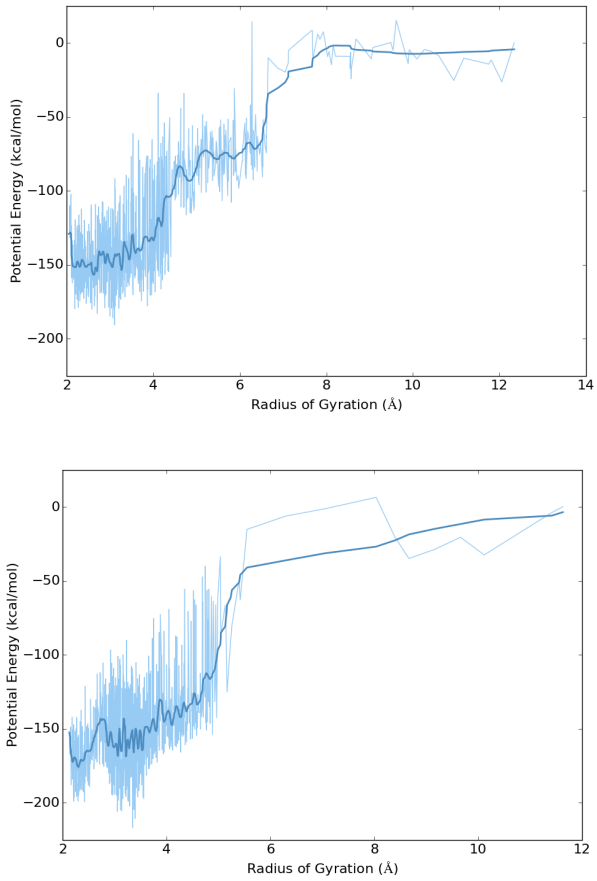


Figure 4.13: Potential energy values plotted against the radius of gyration for trajectory 1 (left) and 2 (right). A smoothing filter has been applied to produce the dark blue line and extract an average trend in the raw (light blue) data.

large standard deviation of the DFT results due to limited sampling is taken into account, the HSE06 potential energies are consistent with the SMRFF potential energies.

The DFT results show large fluctuations in the system energy as a function of the reaction coordinate due to the fact that they are a finite set of samples from a trajectory with likewise high variability in potential energy. In the limit of full sampling, the DFT results should converge to a consistent trajectory of lowest-

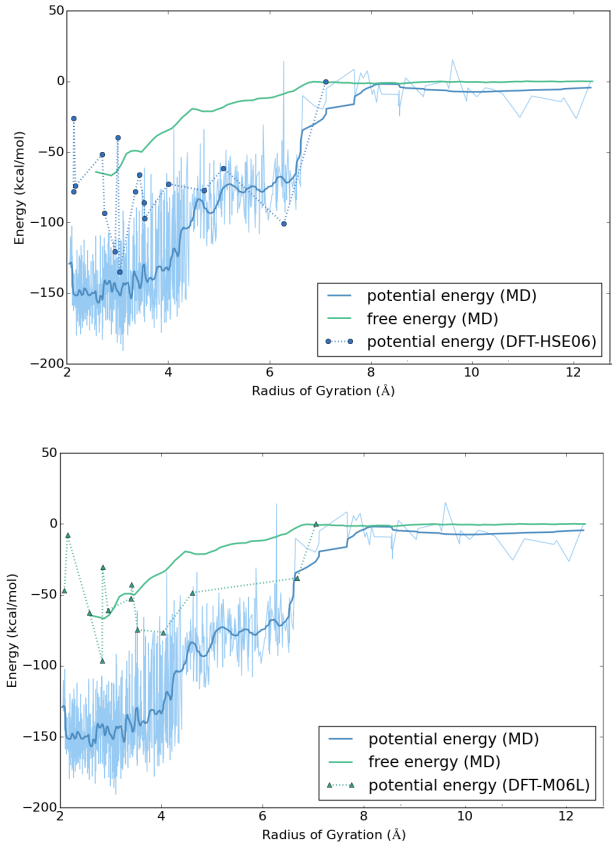


Figure 4.14: Comparison of energy calculations performed *via* Molecular Dynamics (MD) using the metadynamics method and those performed using DFT for the first simulated trajectory. DFT calculations performed using the HSE06 functional are displayed in the top figure, and those using the M06L functional in the lower figure.

energy states at each reaction coordinate. Another sampling-related observation is the difference in energy between states in the locality of those analyzed here, for radii of gyration ≈ 2-7 Angstrom, and the state of infinite separation between the quantum dots. The DFT-calculated energy of two non-interacting dots, at an assumed separation of infinity, ranged from 65 kcal/mol to 138 kcal/mol lower than that of the configuration with the largest radius of gyration from those sampled. The discrepancy in these local *vs.* global energies can, in part, be

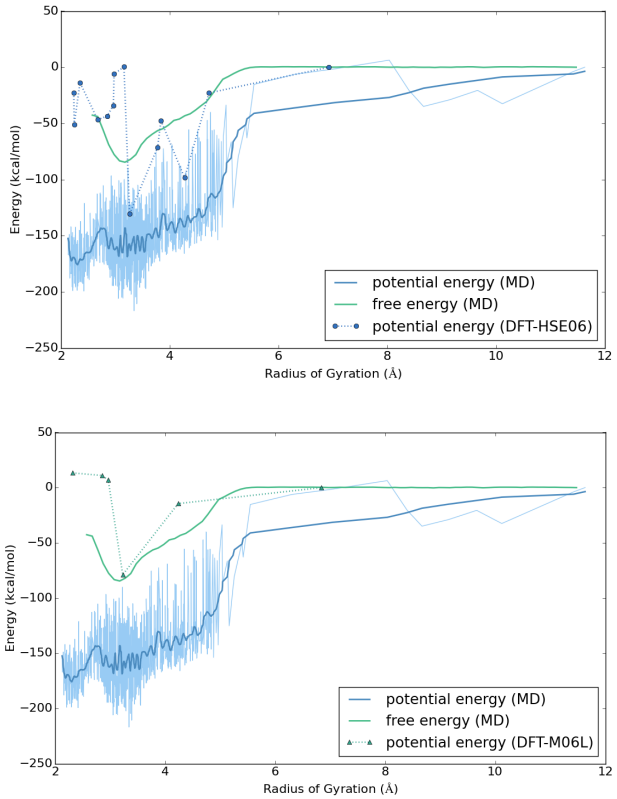


Figure 4.15: Comparison of energy calculations performed *via* molecular dynamics (MD) using the metadynamics method and those performed using DFT for the second simulated trajectory. Key as in Figure 12.

attributed to the sensitivity of the energy to small atomic displacements; a large energy contribution may result from a few geometrically unfavorable atomic positions that hinder the system's optimization to the global energy minimum. Increased sampling of system configurations, as well as an investigation of the behavior of larger dot sizes, may further help complete the picture of potential energy comparisons for simulated interactions like those presented here.

4.3 Conclusions

We have developed a new interatomic/intermolecular model, which we named SMRFF, that is suitable for use in a reactive system. For a test case of PbS, we demonstrated that a SMRFF model was effective at predicting structures and energies across a range of very small nanocrystals, which may be considered as "proto"-dots existing in the earliest stages of nucleation. Encouragingly, the errors in energy are comparable to differences between the predictions of different DFT functionals. Thus, the predictions of our new semi-empirical model lie within the tolerances accessible from a DFT-based method of parameterization. We also favorably compared SMRFF predictions of a system involving the reaction between two dots of size 1 with a DFT prediction of the same reaction although, interestingly, we found the DFT results to be unexpectedly noisy.

A notable property of our chosen potential is its capacity to accommodate a variable number of parameters. Lennard-Jones interactions and Morse interactions can easily be substituted for one another within the same parameter set, allowing coarser or finer representations of particular interactions. Impressively for a pairwise potential, our model predicts the nature of the complex/complex, core/complex, and core/core interactions over a range of surface coverage densities and quantum dot sizes. As a result of these encouraging results, we believe that this model and parameterization technique may be more broadly applicable to other similar systems, certainly other lead and cadmium chalcogenides.

Our approach takes advantage of the flexibility of the Morse potential paired with relevant OPLS potentials and electrostatic interactions. We have thor-

oughly discussed techniques that may be implemented for refining the parameterization process, particularly for many-parameter functions. A key advantage of our new model is its simplicity and the relative ease with which non-specialists can parameterize their own metal-organic system, tailored to their own specialized needs. This allows it to fill a niche above OPLS-like models and below REAX-like models in terms of applicability to different types of metalorganic systems.

SMRFF is also between OPLS and REAX in terms of performance, as depicted schematically in Figure 4.16. The computational expense of Molecular Dynamics simulations scales inversely with the allowable MD timestep, and while OPLS tolerates MD timesteps up to 2.0 femtoseconds, a timescale determined by the fast oscillation of hydrogen single bonds, REAX requires timesteps of no more than 0.5 femtoseconds [85, 53]. SMRFF hydrogen is identical to OPLS hydrogen, so as long as hydrogen oscillations are the constraint, SMRFF and OPLS use the same timesteps. In terms of calculations per timestep, OPLS and SMRFF are also similar, with REAX being an order of magnitude more costly. OPLS is the simplest with its polynomials and trigonometric functions; SMRFF is slightly more costly, adding one exponential function each for certain pairs of atoms. REAX is much more costly than either, summing thirteen energy terms each for every pair of atoms in the system, most of which terms contain multiple exponential or power functions [53]s. These considerations will depend on the implementation, but in general REAX will be slower than OPLS by a factor of ten or more, while SMRFF can be expected to perform within a factor of two of OPLS.

The new reactive SMRFF potential effectively models metal-organic PbS quan-

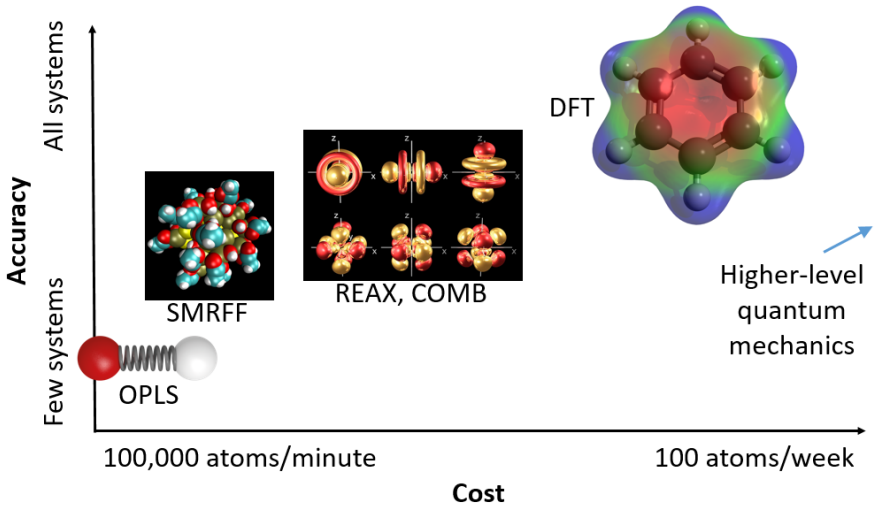


Figure 4.16: Schematic representation of the relative ordering of different levels of theory for all-atom simulation, including SMRFF. SMRFF is much faster than REAX or COMB with little sacrifice of accuracy. It is, however, transferable to fewer types of systems because of the limitations of its functional form, *e.g.*, no reactive angle dependent terms.

tum dots, allowing future studies to be made of the system's reaction dynamics. The existence of this model will help us to predict the evolution and resulting size distribution of PbS quantum dots on larger scales, thus providing a sound fundamental basis of knowledge about the nucleation and early-stage growth properties of lead chalcogenides. This may pave the way to their manufacture on a larger, manufacturing scale.

CHAPTER 5

RATIONAL SOLVENT ENGINEERING FOR HYBRID ORGANIC-INORGANIC PEROVSKITE (HOIP) PRECURSOR SOLUTIONS

All of the most effective HOIP synthesis techniques rely on solution-processing [35, 96]. In the "one-step" synthesis process, methylammonium halide (CH_3NH_3X) is added to a lead halide (PbX_2) solution, forming colloidal particles which produce a HOIP film as the solvent is dried, or as an anti-solvent is added [38, 96]. In the "two-step" process, by contrast, the PbX_2 solution is dried into a solid PbX_2 film, to which a CH_3NH_3X solution is added [35]. The formation of an intermediate Pb-DMSO solid complex has also been found to be helpful [35], as has post-treatment solvent annealing [97, 98]. In all of these processes, the quality of the result depends on the interaction of the Pb compounds with the solvents.

There is an acknowledged demand in the literature for studies that describe the structure of precursor HOIP solutions and how they affect the thin film HOIP product. [38, 36] For example, the drying process which creates the film is controlled by temperature, time, and vapor pressure. The ability to change the solvent mixture could provide researchers the ability to engineer the vapor pressure curve to allow a larger processing "window" and easier, safer manufacturing [99]. Later, as the film deposits, it typically encounters a dewetting problem due to mismatch between the solvent and the TiO_2 or other substrate [27]. Rational solvent engineering would allow us to choose a mixture which better matches the substrate while maintaining sufficient PbX_2 solubility. Moreover, some solvents can interact with lead halides to form a solid complex, while

others cannot, and beneficial effects on the film can be created by mixing these two types of solvents [96]. Other solvents, such as alcohols [97, 98] and methylamine [100, 101] can be used as post-treatments to remove defects and even regenerate aged cells. Finally, control of solvation allows us improve the stability of the final film by creating enduring complexes with acidic linkers and even water itself [27, 102, 103].

In this work, we present the first joint theoretical and experimental study of the solvent effects which make solution-processed perovskites possible. Our calculations accurately rank the experimental solubilities of Pbx_2 in important solvents from first principles, and provide a fast screening tool to quickly predict the effectiveness of any future solvent candidate. This screening uses a computationally efficient analysis of the electron density around the solvent's atoms, linking lab scale results with the fundamental properties of individual molecules. The accompanying experimental work, by our collaborators Dr. Petr Khlyabich and Prof. Lynn Loo at Princeton, supports our theoretical predictions. We also link our screening tools with the broader literature of metal coordination complexes, strengthening the theoretical and practical foundation we provide for solvent engineering in HOIP synthesis.

5.0.1 Previous Theoretical Approaches

There are many solvent properties which might be able to explain the observed lead ion solubilities and complexation properties of the solvents used in HOIP synthesis. These properties include Hansen solubility parameters, dielectric constant, dipole moment, atomic charges, molecular mass, molecular volume, *etc.* One popular concept currently advanced in the HOIP-related literature is to

rank the efficacy of solvents by their relative polarity E_T^N [37], an experimentally-determined polarity scale [104]. As we will show in the Results section, none of these proposals are reliable, including polarity.

Instead, we propose a new predictor. Our hypothesis is that, since the solubility is believed to be dominated by complexation, and complexation is dominated by dative bonding, the solubility should be determined by the solvent's electronic state. A simple way of quantifying the electronic state is the Mayer bond order of the solvent's most electronegative atom (here, N or O). The Mayer bond order partitions the electron density so as to quantifies the degree of bonding on a scale where a perfectly satisfied double bond is 2.0, a triple bond is 3.0, and so on [105]. The idea that low Mayer bond order correlates with increased ability for dative bonding in passivating molecules has been used previously for polyoxometalates of molybdenum [106]. The concept of reactivity created by low bond order dates back to 1931 and the seminal work of Mulliken [107], who called it "bonding power". Solvents which exhibit a high "bonding power" will also be Lewis bases [37, 36].

Experiment and *ab initio* calculations agree that Pb (II) is apt to form complexes with polar ligands [108]. Pb (II) complexes with carboxylates are well characterized; in contrast, lead complexes formed in the solutions used in HOIP syntheses have not been characterized at all [108]. A key distinction between different types of lead complexes is whether the distribution of ligands is uniform across the Pb^{2+} surface (holodirected) or shows significant gaps (hemidirected). Valence Shell Electron-Pair Repulsion (VSEPR) theory applied to Pb-O complexes yields the coordination polyhedra shown in Figure 5.1.

All known Pb(II) compounds with coordination numbers above 8 are holodi-

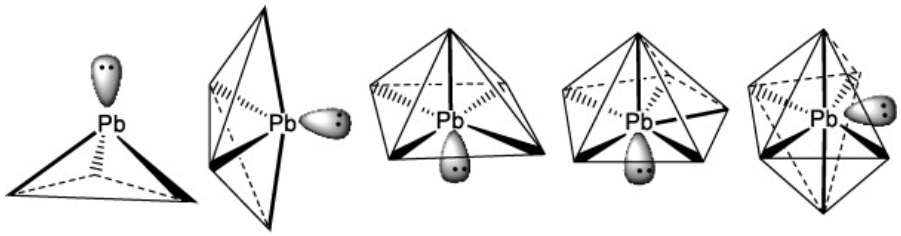


Figure 5.1: Pb^{2+} complex structures suggested by VSEPR theory for different numbers of complexing groups (3-8 from left to right) [108]. Our highly optimized complexes show many more ligands and less order than these regular polyhedra. Figure from Davidovich *et al.*, [108], used with authors' permission

rected, while those with coordination numbers of 8 or below may be either holodirected or hemidirected [108]. Yoon *et al.* proposed that Pb^{2+} would interact with solvents at six coordination sites [36], implying an either holodirected or hemidirected structure. However, we will show below that it is energetically preferred for Pb^{2+} to interact with eight, or even nine, solvent molecules, making it holodirected.

Solvent additives, such as a stoichiometric excess of CH_3NH_3I, have also been used to boost the poor solubility of the lead halides (PbX_2) [38]. In the literature, it has been stated that the effect of CH_3NH_3X on PbX_2 solutions is fully explained by the analagous effect of KI on PbX_2 soltions [109], which slightly contradict the common ion effect (Figure 5.2). However, the effect of CH_3NH_3X is to make the total solubility higher than it is in the pure solvent [38], which is quite different from the effect observed for KI which does not make the ultimate solubility of PbX_2 higher than it is in the pure solvent KI [110]. Therefore this issue must be regarded as still unresolved.

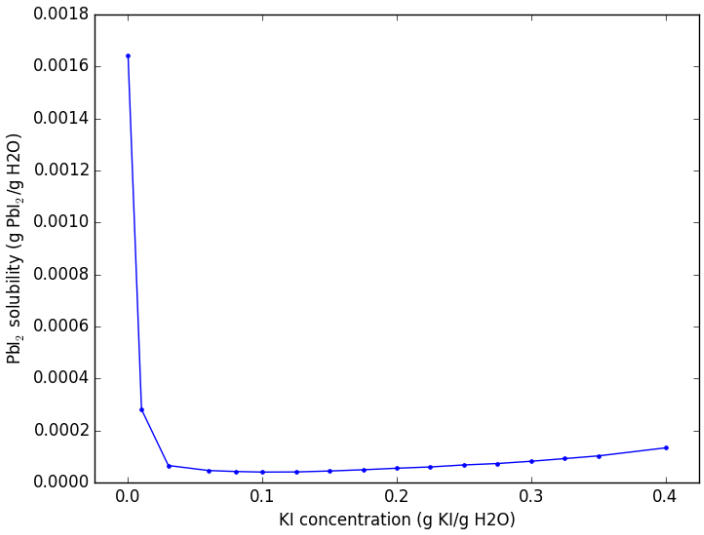


Figure 5.2: The anti-common-ion effect observed for PbI_2 and KI, shown by the upward slope of the right side of the curve. This effect does not fully explain the effect of CH_3NH_3X on PbX_2 solutions, in which the solubility with CH_3NH_3X is *higher* than the solubility in the pure solvent, which would correspond to the rightmost point being higher than the leftmost point. Data from Lanford and Kiehl, 1941[110]

5.1 Experimental

We analyzed the behavior of seven solvents: dimethyl sulfoxide (DMSO), dimethyl formamide (DMF), n-methyl pyrrolidone (NMP), gamma-butyrolactone (GBL), acetonitrile (ACN), acetone (ACET), and methacrolein (METH). Five of these (DMSO, DMF, GBL, and ACN) have all been used as solvents in previous experimental studies of HOIP materials [96, 111]. We added two more solvents, acetone and methacrolein, although they have not been used experimentally for HOIP synthesis. Their role here was to act as "structural controls," since they have molecular structures similar to DMSO and DMF, respectively, but differ as explained by Figures 5.3 and 5.4. Comparison of the solution behavior of ace-

tone and methacrolein to their structural analogs, DMSO and DMF, will help us to uncover the all-important connection between structure and properties for the Pb^{2+}/solvent system.

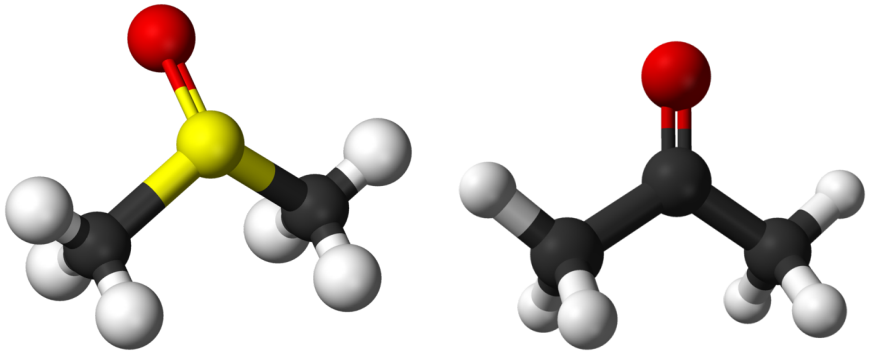


Figure 5.3: Comparison of DMSO (left) and acetone (right). They are analogous in structure *and* isoelectronic, differing in that DMSO is centered on a hypervalent sulfur atom, while acetone is centered on an octet-satisfying carbon atom. Colors: S=yellow, O=red, C=black, H=white

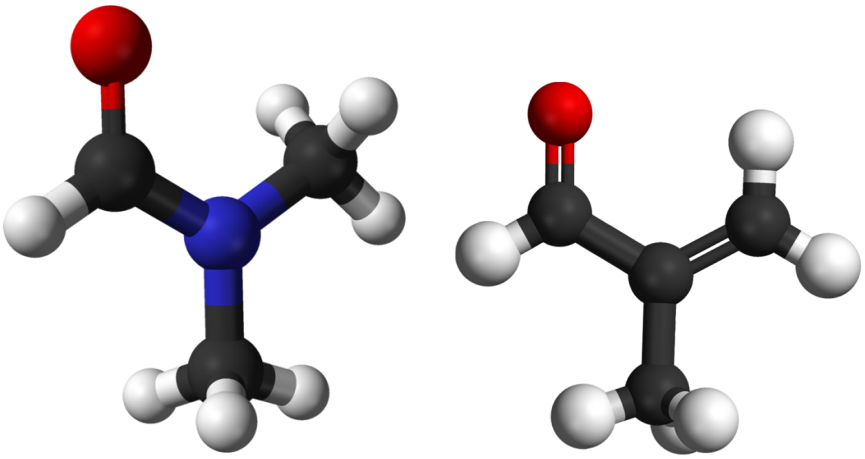


Figure 5.4: Comparison of DMF (left) and methacrolein (right) structures. They are analogous in structure but *not* isoelectronic, since methacrolein has an additional double bond which conjugates with its C=O double bond. Colors: O=red, N=blue, C=black, H=white

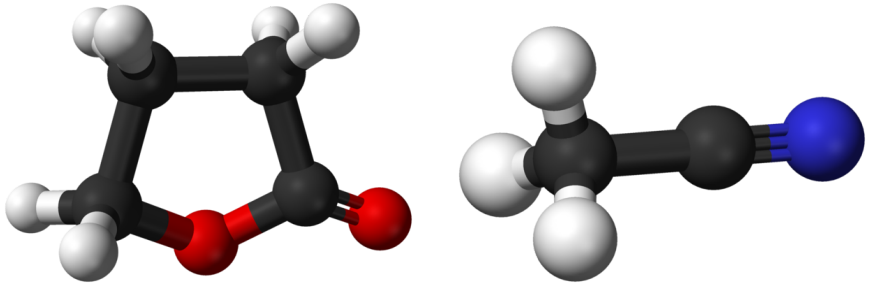


Figure 5.5: Comparison of two less common solvents for HOIP synthesis, GBL (left) and acetonitrile (right). They differ from the pattern set by DMSO and DMF in having either two oxygen atoms (GBL) or no oxygen atoms (acetonitrile). As will be shown, the electronegative atoms in the solvent play a critical role. Colors: O=red, N=blue, C=black, H=white.

5.1.1 *In silico* experiments

The solubility of PbX_2 in polar solvents is closely related to the enthalpy of solvation, ΔH_{solv}, defined thermodynamically as the enthalpy released when a mole of the solute is moved from vacuum to the solvent at infinite dilution. We began by calculating solvated and unsolvated structures for PbX_2, PbX^+, Pb^{2+}, and X^- for three choices of the halide ion, X, as Cl, Br, and I using accurate *ab initio* calculations, described below.

We observed that $\Delta H_{solv:Pb^{2+}}$ was quite predictive of ΔH_{solv} for all the other Pb-containing solutes. This is reasonable because even non-halide Pb salts, like lead acetate and lead nitrate, produce lead halide perovskites when exposed to halides in solution. [112] This suggests that Pb^{2+} becomes totally separated from its original counterions during the perovskite synthesis. It also had the practical advantage that it allowed us to focus on the Pb^{2+} system alone and perform accurate quantum mechanical calculations of $\Delta H_{solv:Pb^{2+}}$.

For each solvent, we created fifty different sample [Pb^{2+} + solvent] geometries, and fifty pure solvent cluster geometries, using simulated annealing and *ab initio* optimization as described below. This sampling is necessary because the system has many degrees of freedom and no optimization of a single posited structure will reliably find its global optimum geometry. Once we found accurate geometries for both the pure solvents and the solutions, we calculated $\Delta H_{solv:Pb^{2+}}$ for each solution as:

$$\Delta H_{solv:Pb^{2+}} = E_{[Pb^{2+}+solvent]} - E_{[solvent]} - E_{[Pb^{2+}]} \tag{5.1}$$

where $E_{[Pb^{2+}+solvent]}$ is the energy of a Pb^{2+} and solvent complex, $E_{[solvent]}$ is the energy of a pure solvent cluster, and $E_{[Pb^{2+}]}$ is the energy of an isolated Pb^{2+} ion in vacuum.

We used a "Jacob's Ladder" approach to generate the structures, moving through increasingly accurate (and computationally expensive) approaches from classical force fields to pure Density Functional Theory (DFT) to more accurate double-hybrid Density Functional Theory [113]. We created initial geometries for the coordination complexes *via* simulated annealing with the semi-empirical OPLS force field [85]. Using those as a starting point, we then optimized the initial geometries using B97-D3[114] / def2-TZVP [74] to obtain more accurate structures [43]. Using these optimized geometries, we evaluated the energy and electronic states using the double-hybrid DFT functional PWPB95 [113], with the very large ("quadruple zeta") basis set, def2-QZVPP(-d,-f) [74].

In order to explore the more fundamental causes of the solubility differences that we discovered for different solvents, we calculated a series of electronic

properties for each solvent: the dipole moment, Mulliken and Loedwin atomic charges, the occupation of the orbitals for the most electronegative atom, and the Mayer bond order for the most polar bond. We calculated these electronic properties using "unrelaxed" electron densities, such that the MP2 correction is not applied to the electronic properties. The Mayer bond order appears to be the most useful property, as will be discussed in Section 5.2.4.

We used the free quantum chemistry package Orca[62] for the DFT geometry optimizations and energy evaluations. We applied the Coulomb-fitting approximation for Coulomb integrals and the RIJCOSX approximation for Hartree-Fock exchange integrals, as described by Weigend [74]. For all the systems containing solvent molecules, but not the Pb^{2+} in vacuum, we added *ab initio* implicit solvation *via* the Solvation Model Density (SMD) method, using the bulk dielectric constant of each solvent [75]. This adds the effect of additional shells of solvent beyond the first, explicitly represented solvent molecules which complex with the solute. We performed the Molecular Dynamics simulations in LAMMPS[95] with starting geometries created in Packmol[115], each with a different pseudo-random seed. We used standard OPLS 2005 parameters for all solvent molecules and treated Pb^{2+} (which is not parameterized in OPLS 2005) as Ba^{2+}, relying on the subsequent DFT optimizations to remove the slight overcrowding of the Pb^{2+} which results from Ba^{2+}'s smaller van der Waals radius.

We also studied the impact of solvent blends, considering mixtures of DMSO:DMF, DMSO:acetone, and DMF:acetone, as clusters around Pb^{2+} with six solvent molecules in total for each cluster for computational expedience. For these mixed solvent studies, we did not perform Molecular Dynamics sampling because of the likelihood that the solvent ratio in the cluster would be different from that in the

bulk mixture. We also used isolated solvent molecules in implicit solvent for the baseline energy comparison, instead of optimized solvent clusters, once again to improve the speed of the calculation. Our standard of comparison for these studies was therefore $\Delta H_{cluster:Pb^{2+}}$, slightly different from $\Delta H_{solv:Pb^{2+}}$, calculated as:

$$\Delta H_{cluster:Pb^{2+}} = E_{[Pb^{2+}+solventsA,B]} - n_A E_{[solventA]} - (6 - n_A) E_{[solventB]} - E_{[Pb^{2+}]} \quad (5.2)$$

where $E_{[Pb^{2+}+solventsA,B]}$ is the energy of the complex, n_A is the number of molecules of solvent A in the cluster, $E_{[solventA]}$ is the energy of a single molecule of solvent A, $E_{[solventB]}$ is the energy of a single molecule of solvent B, and $E_{[Pb^{2+}]}$ is the energy of an isolated Pb^{2+} ion in vacuum.

5.1.2 *In vitro* experiments

Our colleagues Dr. Petr Khlyabich and Prof. Lynn Loo at Princeton performed solubility measurements for DMSO, DMF, GBL and ACET. These measurements were using a glovebox under inert atmosphere to avoid contamination, especially from water vapor. The solvents and PbX_2 salts were not exposed to air prior to the measurements. This is especially important for DMSO since it is hygroscopic [99]. Lead halide salts with 99.999% purity, and solvents with purities above 99.5%, were purchased from Sigma Aldrich and used without further purification. No solubility-enhancing additives, such as methylammonium halide, were used in these solutions.

Dr. Khlyabich performed the measurements at room temperature, approximately 25°C. Each lead salt was added separately to a vial with each solvent and stirred for 15 min. If no precipitate was observed, more salt was added to the solution. The solubility of the lead halide salt was estimated when the precipitate was observed after 30 min of continuous stirring. The expected experimental error, based on the difference between two trials for each data point, is estimated to be less than 3%.

The property $\Delta H_{solv:PbX2}$ is technically accessible to experiment, using calorimetry to calculate the enthalpy of *solution* for the solid PbX_2 salts, and then adding the salts' lattice energies (since the reference state for the enthalpy of solvation ΔH_{solv} is the solute vacuum rather than in the solid phase). Our collaborators did not have the necessary apparatus to perform these experiments, but as a subject of future work, experimental values for $\Delta H_{solv:Pb^{2+}}$ would be helpful.

5.2 Results and Discussion

5.2.1 Structures of Coordination Complexes

Our calculations agree with experimental supposition that the optimum state for Pb^{2+}/solvent clusters is a coordination complex. However, we find that the preferred Pb^{2+} complex, in the presence of abundant solvent, contains eight or nine solvent molecules, rather than the six molecules suggested by simple Valence Shell Electron-Pair Repulsion (VSEPR) theory [108] (Figure 5.1) and proposed by *Yoon et al.* [36]. This is understandable because VSEPR theory for metal complexes becomes less accurate with increasing atomic number [36].

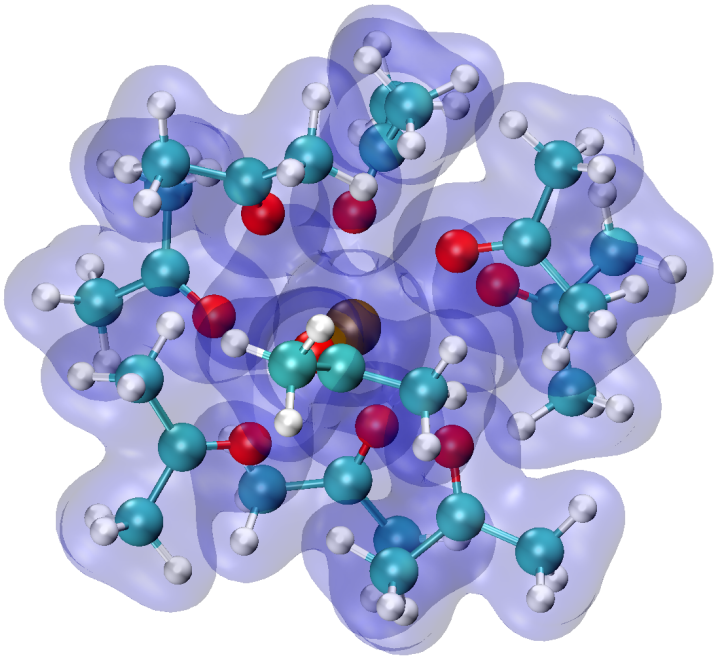


Figure 5.6: Electron density predicted by our *ab initio* calculations surrounding a Pb^{2+}/acetone cluster with nine coordinated oxygen atoms, which we found to be the energetically preferred number of solvent neighbors. Color key: Pb = brown, O = red, C = cyan, H = white. The blue shading represents the electron density at the 0.005 $e^{-}/^{3}$ isosurface

Similarities are immediately apparent between the coordination complexes created by the different solvents. The number of electronegative atoms bonded to each Pb^{2+} ion in the DFT-optimized structures is always either 8 or 9 across all the solvents we studied, without regard to their solubility, and regardless even of whether that central atom is oxygen or nitrogen. This could suggest that the structure is simply the result of packing the electronegative atoms around the positive Pb^{2+} ion. However, the ΔH_{solv} results tell a very different story. Despite their geometric similarities, different solvent complexes produce very different solubilities, as described in the next section.

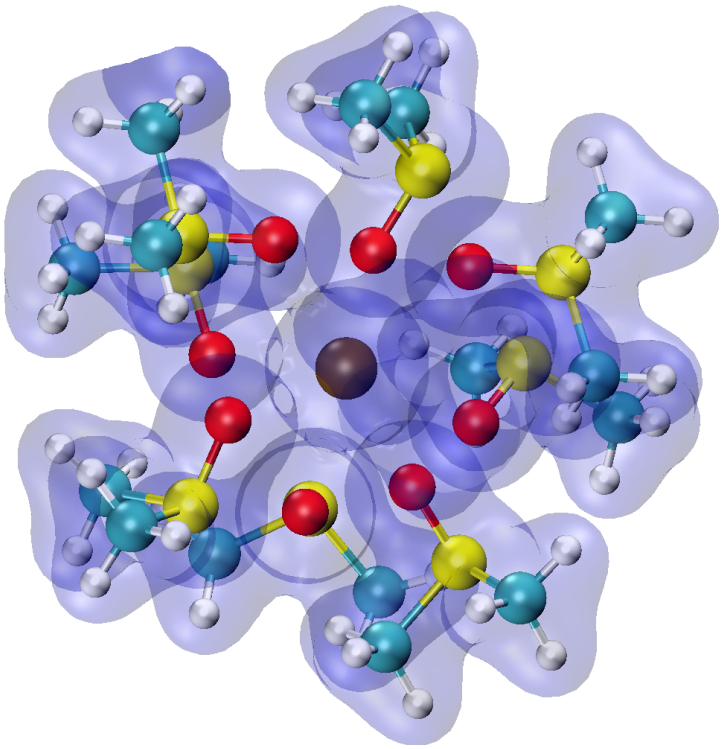


Figure 5.7: Electron density predicted by our *ab initio* calculations surrounding a Pb^{2+}/DMSO cluster with eight coordinated oxygen atoms, the energetically preferred number of neighbors for this solvent. Color key: Pb = brown, S = yellow, O = red, C = cyan, H = white. The blue shading represents the electron density at the 0.005 $e^{-}/^{3}$ isosurface

5.2.2 Computation of ΔH_{solv}

The results of our initial studies on PbX_2, PbX^+, Pb^{2+}, and X^- are shown in Figure 5.8 as a plot of ΔH_{solv} for seven solvents, lead ions, and different halides. The fact that the lines for a given solvent rarely cross each other shows that the relative ranking of $\Delta H_{solv:PbX_n}$ is stable regardless of which PbX_n complex is being used. This motivated our selection, mentioned above, of the simple Pb^{2+} system for enhanced sampling (*i.e.*, consideration of 50 different initial configurations).

The more halide ions attached to the Pb^{2+} ion, the lower the value of ΔH_{solv}.

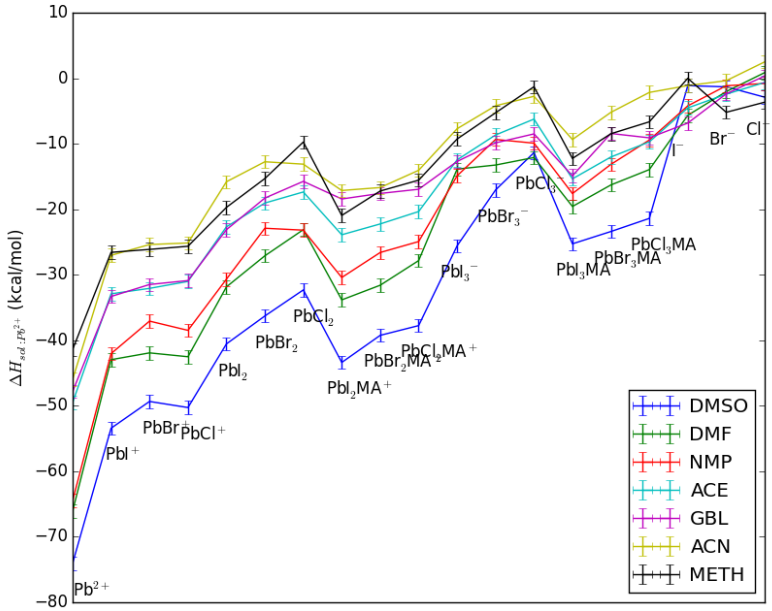


Figure 5.8: *Ab initio* predictions of ΔH_{solv} for PbX_2, PbX^+, Pb^{2+}, and X^- complexes in pure solvents, with $CH_3NH_3^+$ as noted in the labels. Solvent key provided in the inset

This is consistent with the fact that the halide ions serve to passivate the Pb^{2+}. Where present in the solute, the contribution of the halide to the solubility follows the expected periodic trend, with chlorides being least soluble and iodides most soluble.

The effect of methylammonium halide, also shown in Figure 5.8, is striking. $CH_3NH_3^+$ forms hydrogen bonds with up to three X^- ions, as in Figure 5.9. For complexes with DMSO, the addition of methylammonium halide stabilizes the complexes with X−3 by 10 kcal/mol for X=Cl, 6.5 kcal/mol for X=Br, and 0.0 for X=I. This makes sense because iodine is known not to form strong hydrogen bonds. This computational prediction helps to explain how methylammonium halide acts as such an effective solubility enhancer: its ability to form a com-

plex with up to three halide ions at a time stabilizes agglomerations of Cl^- and Br^-. More fundamentally, the methylammonium triple-halide interaction helps to explain the HOIP synthesis, showing how CH_3NH_3X acts as a chaperone to bring together the other reactants and form the basis for the perovskite lattice, as shown by the perovskite "dimer" in Figure 5.10. A study of the dimerization enthalpies along with activation energies E_a, as a function of the choice of solvents, halides, and cations, would be extremely useful as a piece of future work, since it would allow prediction of the kinetics of the perovskite synthesis. A classical HOIP force field would be helpful in such a study to improve the sampling of the complex geometries.

Figure 5.11 shows the motivation for sampling many randomized starting geometries for each solvent cluster. If only one sample were used per solvent, the energy might be 5 or 10 kcal/mol higher than the true optimum energy, misordering some of the results. ACN has the smallest range of energies and GBL has the largest, which directly corresponds to the size and complexity of the molecules. ACN clusters have fewer possible energy states because ACN is small and linear, almost radially symmetric, while GBL is an irregular ring and not symmetric.

5.2.3 Experimental Solubilities

Table 5.1 shows the experimental solubilities of Pb salts in the pure solvents.

While PbI_2 and $PbBr_2$ show similar solubilities across all studied solvents, $PbCl_2$ demonstrates significantly reduced solubility in comparison to the other halides, and shows fair solubility only in DMSO. In general, for all studied sys-

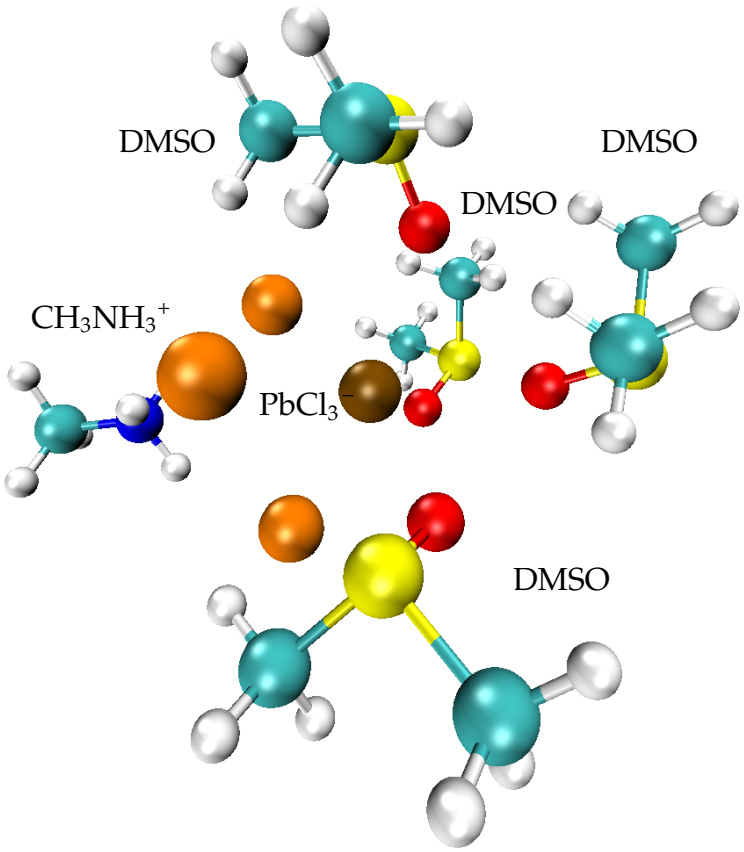


Figure 5.9: Lead chloride perovskite "monomer" passivated by DMSO solvent. The Cl ions are stabilized by hydrogen bonds with $CH_3NH_3^+$. Color key: Pb=brown, Cl=orange, S=yellow, O=red, N=blue, C=cyan, H=white

Table 5.1: Experimental solubilities of lead halide ions for the pure solvents at room temperature

	DMSO	DMF	GBL	ACET
PbI_2	600 mg/ml	450 mg/ml	<4 mg/ml	<5 mg/ml
$PbBr_2$	560 mg/ml	350 mg/ml	<4 mg/ml	<4 mg/ml
$PbCl_2$	310 mg/ml	17 mg/ml	<2 mg/ml	<2 mg/ml

tems, DMSO demonstrated the highest solubility, followed by DMF, ACET and GBL. These data show similar trends to the theoretical predictions presented above. The experimental and theoretical observations in this study thus explain

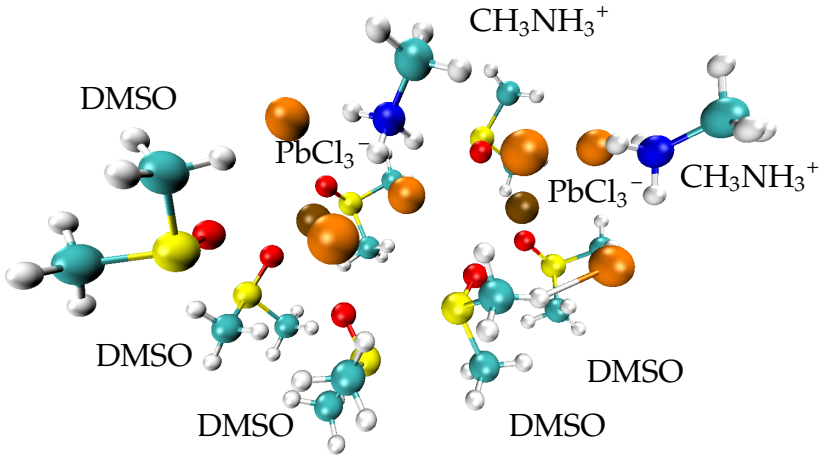


Figure 5.10: *Ab initio* predictions of the structure of a (HOIP) perovskite "dimer" in a coordination complex with DMSO solvent. The methylammonium ions are clearly visible with their distinctive dark blue N atoms. Color key: Pb=brown, Cl=orange, S=yellow, O=red, N=blue, C=cyan, H=white

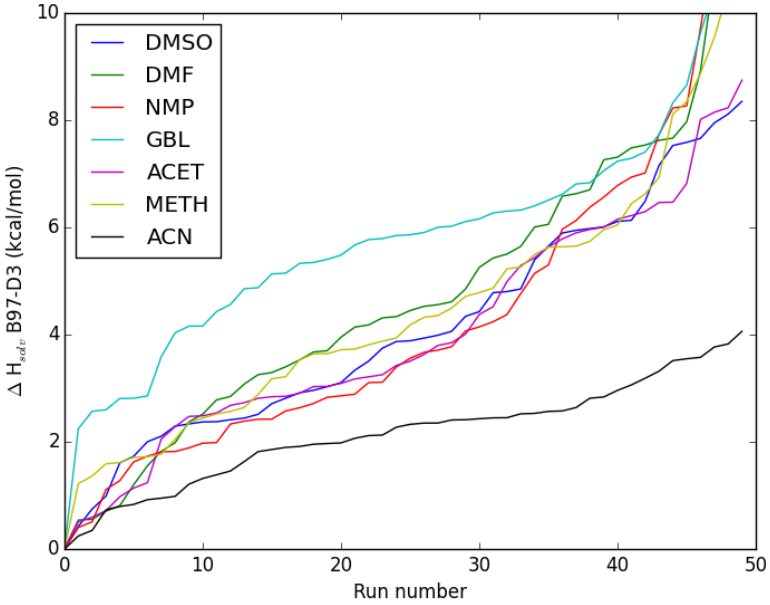


Figure 5.11: The energies of the sampled states of the Pb^{2+}/solvent complexes, with 50 samples for each solvent. The baseline for each curve is the lowest-energy state for that complex

the relative ranking of solvent effectiveness to improve solubility as used for methylammonium lead halide perovskite processing. The vast majority of active layers are currently processed using either DMSO or DMF. Moreover, it explains the selection of GBL as an additive frequently used for the active layer fabrication, since GBL can facilitate aggregation in solution which, in thin films, can serve as nucleation centers during thin-film formation.

It is interesting to note that the mole ratio of Pb:solvent in the DMSO and DMF solutions is large, as high as 1:9 for $PbBr_2$ in DMSO. With solubility enhancers such as methylammonium halide, the mole ratio would be even greater. Solute-solute interactions, or at least interactions between adjacent Pb/solute complexes, will therefore be significant since each complex prefers to hold as many as nine solvent molecules. Infinite dilution, as in the definition of enthalpy of solvation, does not hold strictly for such solutions; a soft network of interacting complexes would be a more accurate picture of the system if it were necessary to simulate such a solution with complete veracity.

Interestingly, Table 5.1 shows that pure acetone provides a similar PbX_2 solubility to that of the much more commonly used GBL. The use of acetone as an additive has not been reported in the literature to the best of our knowledge. It appears that GBL is much more enhanced by the presence of methylammonium halide than acetone is, making GBL practical for perovskites despite having a poor solubility for the pure lead halide. Alternatively the problem may be the low boiling point of acetone (56°C, compared to 204°C for GBL), leading to faster evaporation of the solvent during film drying, which could be detrimental to the final film quality. However, based on our theoretical results, acetone is a plausible supplement to GBL and other solvents in mixtures, especially for sol-

vent mixtures where it is desirable to evaporate one solvent more quickly (such as mixed DMSO/acetone). Further experimental studies will be necessary to test this point, along with theoretical studies to determine why the solubility-enhancing effect of methylammonium halide appears to vary significantly with respect to the solvent choice.

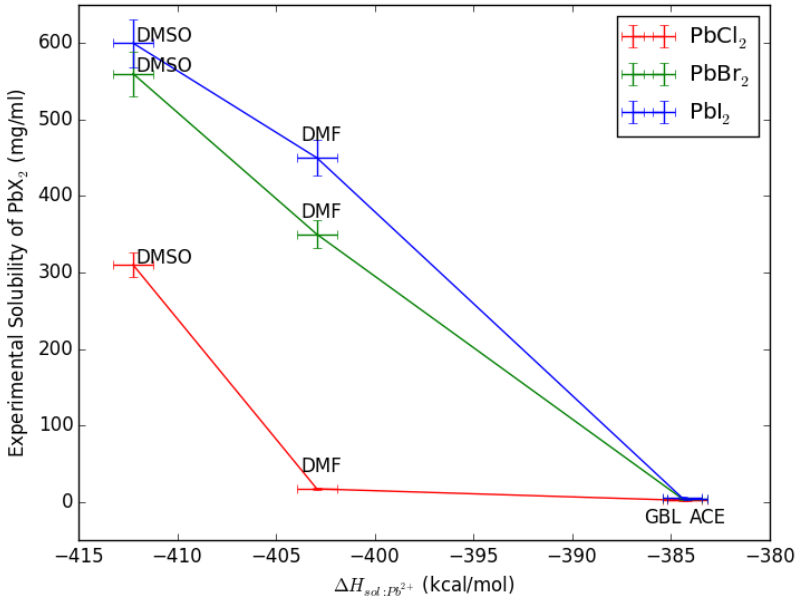


Figure 5.12: Experimental solubility of PbX_2 *vs.* the *ab initio*-derived values of $\Delta H_{solv:Pb^{2+}}$

(the energy values are large because the reference state is vacuum). This relationship shows a linear correlations for the bromide and iodide for DMSO, DMF and GBL (leftmost three points). The result for acetone (the rightmost point) is visually identical to GBL. The correlation for the chloride is flat for much of the range (*i.e.*, for DMF, GBL and acetone) because $PbCl_2$ is a much poorer solute - only DMSO is notably effective. Color key for the halides is given in the inset

5.2.4 Fast *ab initio* Prediction of Solubility

While the DFT results described above have provided invaluable information that allowed us to determine the preferred structures for lead halide ions in the presence of solvent and chaperone ions like methylammonium, and produce solubilities that are in agreement with our experimental results, they are very costly in computational time. Each of rhese *ab initio* calculations required about 200 hours of CPU time per double-hybrid DFT analysis of each cluster (the most expensive step), totaling about 100,000 CPU-hours (10 CPU-years) for all the samples of all the cluster types described in this paper. This provides considerable incentive to look for a simpler chemical correlation which is predictive enough to screen good solvents from poor ones for future experimental studies, without the need to undertake costly *ab initio studies*.

We determined that no reliable correlation exists between solubility and common solvent properties such as the dielectric constant, molar mass, molar volume, or Hansen solubility parameters, or relative polarity (see Table 5.2). Each one has significant discrepancies within our set of solvents: for example the relative polarity, used to compare DMSO and DMF by Ahn *et al.* [37], ranks GBL above DMF and acetonitrile above any other solvent including DMSO [104]. These properties are clearly insufficient to explain the chemical origin that differentiates among the solvents.

Our *ab initio* calculations of the electronic-state properties of each solvent are shown in Table 5.3. These properties include the magnitude of the dipole moment and various measures of the solvent's most electronegative atom: its Mayer bond order, p-orbital occupation, Loedwin atomic charge, and Mulliken atomic charge. We observed a strong, albeit not always linear, correlation be-

Table 5.2: Common empirical properties of the solvents. We find that none of these properties are effective at ranking all the solvents considered here

	Dielectric constant	Molar mass (amu)	Molar volume (cm^3)	Hansen δ_P ($MPa^{1/2}$)[116]	Relative polarity E_T^N [104]
DMSO	46.7	78.1	71.3	16.4	0.444
DMF	36.7	73.1	77.4	13.7	0.386
NMP	32.2	99.1	96.5	12.3	0.355
GBL	40.2	86.1	76.2	16.6[117]	0.420
ACET	20.7	58.1	73.5	10.4	0.355
METH	10.9	70.1	86.7	-	-
ACN	37.5	41.1	52.6	18.0	0.460

tween ΔH_{solv} and the Mayer bond order of the solvent's oxygen atom, and also between the experimental solubility of PbX_2 and our estimate of the Mayer bond order (Figure 5.13). In order to use a common scale for both oxygen and nitrogen, we present it as a "Mayer bond unsaturation," consisting of the formal bond order (2.0 for oxygen, 3.0 for nitrogen) minus the calculated Mayer bond order.

We find that the Mayer bond unsaturation is the best measure we studied to capture the effectiveness of the seven solvents that we studied, unlike other empirical properties such as the Hansen polarity parameter (GBL alone is more effective than its primary oxygen's bond order would suggest, perhaps because GBL contains a second oxygen atom). The use of electronic properties as a screening method is especially useful because the electron density of an arbitrary solvent molecule can be calculated in minutes on a personal computer using free software, whereas empirical data for a given solvent candidate may not be available. This is especially true for compounds not currently used as

solvent choices for perovskite synthesis in the lab.

Table 5.3: Electronic properties of solvent molecules: note that the Mayer bond unsaturation and p-orbital occupation of the solvent's electronegative atom are predictive of the experimental solubility ranking, while the dipole magnitude, Loedwin atomic charge, and Mulliken atomic charge are not

Solvent	Bond unsaturation	p-orbital occupation	Dipole magnitude (Debye)	Loedwin charge	M
DMSO	0.50	4.73	5.75	-0.32	
DMF	0.12	4.44	5.37	0.15	
NMP	0.10	4.44	5.46	0.16	
GBL	0.01	4.24	6.28	0.21	
ACET	0.03	4.36	4.09	0.23	
METH	0.03	4.35	4.04	0.24	
ACN	-0.03	3.61	5.66	0.04	

The properties which predict solubility also provide further evidence that the mechanism of dissolution involves dative bonding. In a simple polar solvent model, the dielectric constant and dipole moment should be effective predictors of solubility [118]. Instead, these properties are not predictive for the Pb^{2+}/solvent system, while we find the electronic state of the dative bonding atom (oxygen or nitrogen) in the solvent to be highly predictive. This implies that the availability of dative bonds dominates over simple polarity effects.

Figure 5.13 also shows that, in terms of the solvation of Pb^{2+}, acetone and methacrolein fall well below their structural analogs, DMSO and DMF. Even between solvents which are very similar in geometry, there is a striking difference in Pb^{2+} solubility, the physical basis of which only appears when one considers the solvents' electronic properties.

Table 5.4 shows $\Delta H_{solv:Pb^{2+}}$ and Mayer bond orders at different levels of the-

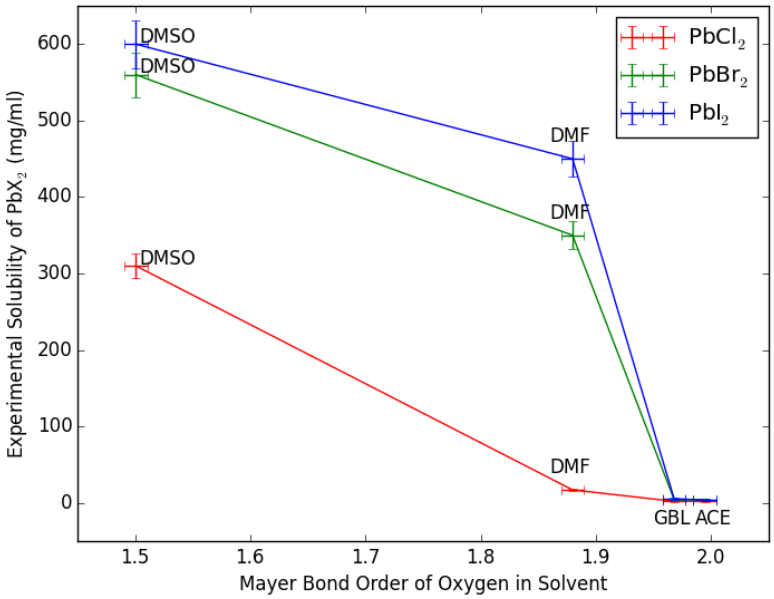


Figure 5.13: Experimental solubility of PbX_2 *vs.* Mayer bond order of the oxygen atom in the solvent. The Mayer bond order follows the same trend as our calculated ΔH_{solv}, making it also an effective predictor of trends in solubility. The error bars reflect expected error in the solubilities and calculated Mayer bond orders. The color key for the halides is shown in the inset

ory. The comparison of pure DFT (B97-D3) and double-hybrid DFT (PWPB95, unrelaxed densities) shows that the values change, but the ordering remains consistent. this shows that the Mayer bond order is an effective ranking prediction tool of solvent quality, both at the expensive PWPB95 level of theory and at the inexpensive B97-D3 level.

5.2.5 Solvent mixtures

Experiments have shown that useful effects appear when HOIPs are synthesized in mixtures of different solvents [96]. In order to test the effect of mix-

Table 5.4: Comparison of $\Delta H_{solv:Pb^{2+}}$, Mayer bond order (MBO) calculated at two different levels of theory, and experimental solubility of PbI_2

	$\Delta H_{solv:Pb^{2+}}$ (kcal/mol)	MBO (PWPB95)	MBO (B97-D3)	PbI_2 sol. (mg/mL)
DMSO	-412	1.50	1.61	600
DMF	-403	1.88	1.94	450
NMP	-401	1.90	2.01	-
ACET	-384	1.97	2.11	5
GBL	-384	1.99	2.10	4
METH	-379	1.97	2.08	-
ACN	-374	3.03	3.01	-

ing solvents, we prepared Pb^{2+} solvated clusters with different compositional blends of DMSO, DMF and acetone, as shown in Figure 5.14, DMSO and DMF being strong solvents commonly used in HOIP synthesis, and acetone being a weak solvent, likely to be used in solvent mixtures only.

The DMSO:acetone and DMF:acetone mixtures show that $\Delta H_{cluster:Pb^{2+}}$, representing mixtures of a strong solvent with a weak solvent, is a simple linear combination of the $\Delta H_{cluster:Pb^{2+}}$ values for the component solvents. The DMSO:DMF mixture is more interesting, with notable deviations from linearity. This is most likely due to the fact that DMSO and DMF are both strong solvents capable of dative bonding, allowing each added solvent molecule to influence the electronic state of the Pb^{2+} ion. Such interactions might affect the other solvent molecules binding to the same ion, making the sum non-linear. Acetone by contrast has a much lower bonding power and thus less impact on the Pb^{2+} ion's electronic state, which may cause its energy interactions to sum linearly per molecule.

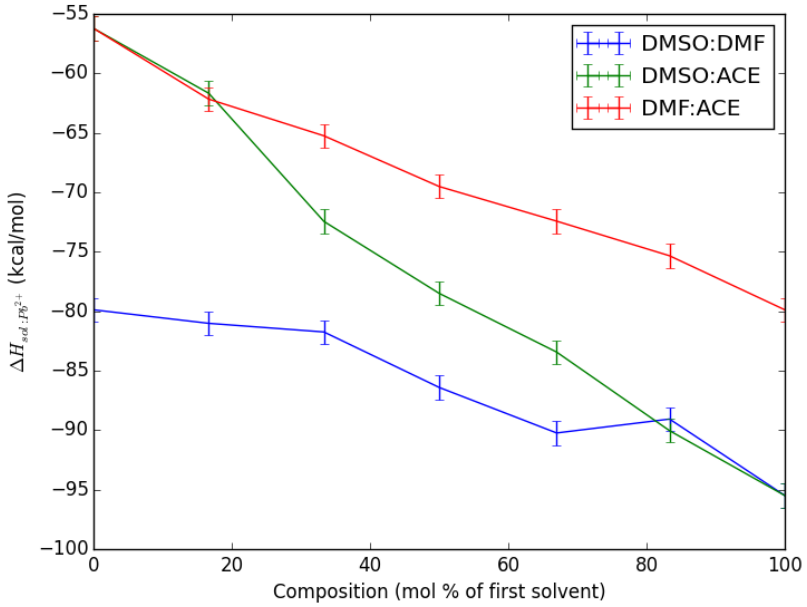


Figure 5.14: $\Delta H_{cluster:Pb^{2+}}$ for clusters of DMSO:DMF, DMSO:acetone, and DMF:acetone. The color key for the solvent blends is shown in the inset

In practice, $\Delta H_{solv:Pb^{2+}}$ for a mixture will be biased toward the stronger solvent because the stronger solvent will increase its local concentration around the solute. Effectively each Pb^{2+} ion will experience a locally elevated concentration of the stronger solvent due to their stronger energy interaction. Therefore DMSO:acetone, DMF:acetone, and DMSO:DMF mixtures are likely to be much more effective solvents than $\Delta H_{cluster:Pb^{2+}}$ for single clusters would suggest.

5.2.6 Screening Alternative Solvents

Using the pure-DFT-derived Mayer bond order of the oxygen as a metric, based on its correlationary role in our preceding work on commonly used solvents for HOIP synthesis, we then screened some common oxygen-containing liquids

for their viability as dative bonding solvents for the perovskite synthesis. This provides a prediction of the effectiveness of as-yet experimentally untested solvents. The results are shown in Table 5.5. The clear standout among the five solvents we studied is nitromethane, an industrial solvent. With a bond order of 1.60, nitromethane's oxygen atoms have a slightly higher bonding power than DMSO's at 1.61. However, nitromethane is not simply a copy of DMSO; it has two oxygen atoms per molecule, which will cause more steric hindrance around the Pb^{2+} ion than any other solvent we have tested. Nitromethane also evaporates much more easily than DMSO, with a boiling point of 100°C, *vs.* 189°C for DMSO, a valuable property for manipulating vapor pressure curves in mixtures. Nitrous acid also has a low Mayer bond order for oxygen, making it another intriguing and untested alternative solvent.

Table 5.5: Proposed dative bonding solvents based on Mayer bond unsaturation of each solvent's double-bonded oxygen atom

Solvent	Bond unsaturation (B97-D3)
Nitromethane	0.40
Nitrous acid	0.12
Thiosulfuric acid	0.07
Hydrophosphorous acid	0.05
Sulfuric acid	0.05
Phosphoric acid	-0.05

5.3 Conclusions

Prior to this work, the only theory of the effectiveness of HOIP solvents available in the literature was that they would be sorted by their empirical relative

polarities [37]. In this work, we have provided clear evidence that the polarity is not effective in determining the solubility or complexation strength in the HOIP synthesis. In every test, high bonding power and Lewis basicity measured by the Mayer bond order proved more important than empirical or theoretical polar qualities, such as the relative polarity, Hansen polar parameter, molecular dipole moment, or dielectric constant.

Our *ab inito* DFT studies, in concert with experimental data, show that we can understand the solubility as an energy-driven complexation process, and that the DFT ranking of $\Delta H_{solv:Pb^{2}+}$ correctly matches the experimental ranking of solubility. Furthermore, the Mayer bond order is an effective ranking tool, whether computed at low- or high- levels of theory (*e.g.*, pure DFT or double-hybrid). Importantly, this approach has the advantage that the calculation of the Mayer bond order is *far* less computationally intensive than energy evaluations of the complexes and, unlike empirical solvent properties, it is readily calculable for any conceivable candidate solvent molecule.

The structures we calculated for the Pb^{2+}/solvent complexes show that they tend to be eight-coordinated, not six-coordinated as previously predicted in the literature [36]. No clear pattern of solute geometry is evident in the optimized, sampled complexes, which makes sense for loosely-bound, eight-coordinated Pb structures [108].

The effectiveness of solvent blends appears to be predictable, with $\Delta H_{cluster:Pb^{2+}}$ showing linear mixing for binary mixtures of weak with strong solvents, and deviations from linearity for mixtures of strong with strong solvents. We suggest that this non-linearity is most likely to occur because "strong" solvents are those which induce dative bonding, and therefore alter the electronic state of the Pb^{2+}

ion. Each new dative bond affects all the other bonded solvent molecules such that the total contribution ceases to be linear.

Our calculations confirm the experimental observation that the most powerful known solvent for the perovskite synthesis is DMSO, followed by DMF. Using the Mayer bond order as a screening tool for other as yet untried solvents, our calculations suggest that nitromethane may also be particularly effective. Acetone is also surprisingly effective; in the absence of the organic halide salt, it is apparently as effective as GBL. Since acetone is a staple of chemical laboratories, it may be a valuable addition to the HOIP synthesis toolkit, especially with the methylammonium salt and/or in mixtures with other solvents such as DMSO. Use of solvent mixtures has another potential benefit: they offer additional degrees of freedom in the process to produce a wide variety of desired solubilities, evaporation rates, and colloidal and solid-state complexation behaviors.

The experimentally observed effect of methylammonium halide as a solubility enhancer is addressed by our cluster geometry calculations, which show that the methylammonium ion stabilizes halide solutes. The NH_3^- group can form up to three hydrogen bonds with halide ions, allowing it to stabilize clusters such as PbX_2 and PbX_3. The effect weakens with increasing atomic number of the halide, with PbI_3^+ having little interaction with NH_3^-, which makes sense because I is less capable of hydrogen bonding than the other halides. Therefore, we cannot yet claim that this effect fully explains the enhanced solubility caused by methylammonium halides, at least for the iodide.

The concentrations observed by experiments are large enough that solute-solute effects may begin to be significant in these solutions, especially once

methylammonium ions are included. Solute-solute interactions of this kind might lead to a solution composed of large, interconnected networks of solvated Pb complexes. Unfortunately, DFT is not computationally efficient enough to simulate such large hypothesized systems, except with supercomputer-class resources. A classical force field would be very useful for this purpose, validated against suitably determined DFT results. Such semi-empirical force fields are several orders of magnitude faster than DFT. Unfortunately, no adequate reactive force field currently exists for studying HOIP solutions.

As the ability to control HOIP synthesis grows in importance, control over the solution processing aspects will only become more vital. Together with temperature and halide composition, solvent choice is one of the few "levers" available with which to continue to improve the efficiency and stability of HOIP solar cells. The theoretical underpinnings we describe, and the fast computational screening techniques we demonstrate, will aid in this process and point the way to more rational solvent engineering for perovskite precursors in the future.

CHAPTER 6

SOLVATION MODELING IN HIGHLY EXOTIC NONPOLAR SOLUTIONS

So far, this work has focused on the practical simulations necessary to understand the synthesis and processing of solar cells in organic solutions. However, organic solution modeling methods have applications far outside of photovoltaics. Among other uses, such simulations allow us to work around missing experimental data by instead starting from first principles and working *up* to the scale of experiments. In the field of solution-processed solar cells, the pieces of missing experimental data arise from the difficulties in measuring certain aspects of the quantum dots and perovskites, such as their atomic structure and reaction mechanisms. There also exist systems, though, which due to physical constraints are troublesome to study experimentally in any aspect whatsoever. Such is the case for the liquid seas of Saturn's largest moon, Titan.

Saturn's moon Titan is the only astronomical body other than Earth that is known to have stable bodies of liquid on its surface. To begin with, the exact composition of these solutions is not known [119]. The current model of their composition, based on radio measurements of the high-latitude lakes in Titan's northern hemisphere, is that the surface liquids are substantially methane, with 5-15% dissolved nitrogen, less than 10% ethane, and other organic molecules in lower concentrations [120]. The light hydrocarbon seas interact with Titan's thick nitrogen-based atmosphere to produce a methane cycle, with methane clouds and methane rain. The solid surface of heavier hydrocarbons is dissolved and eroded by methane flows, forming meandering channels, valleys, and other features familiar on Earth [121].

This system would be complex enough without the added contributions of a

natural chemical factory in Titan's upper atmosphere. Hundreds of kilometers above Titan, the small flux of ultraviolet light from the sun is enough to cause atmospheric methane and nitrogen to form free radicals, producing a range of organic nitrogen-containing compounds. These compounds have been predicted since the 1970s [122], and examples actually observed spectroscopically by the Cassini mission are shown in Table 6.1 [123, p. 176].

Name	Formula	Concentration (ppm)
Hydrogen cyanide	HCN	200-240
Cyanoacetylene	HC_3N	0.8-40
Acrylonitrile	C_2H_3CN	0.3-10
Cyanoallene	C_4H_3N	4
Acetonitrile	CH_3CN	1.5-3
2,4-pentadiynenitrile	HC_5N	1
Propanenitrile	C_2H_5CN	0.15-0.5

Table 6.1: The most common organic nitrogen species observed spectroscopically in Titan's upper atmosphere. Ranges refer to interpretations of the data *via* different ion chemistry models [123, p. 176].

Unlike the hydrocarbons which which compose the bulk of the surface of Titan, these nitrogen compounds are strongly polar, allowing them to possibly resist solvation in Titan's surface fluids [122]. The stability and insolubility of these compounds might cause them to accumulate as grains, pebbles, or boulders on the surface of Titan, while more soluble hydrocarbons were weathered around them by Titan's surface liquids. In addition, insolubility in the liquid is the first step toward compartmentalization, the separation of one part of the liquid from another, a necessary step in the creation of regions of higher concentrations of more advanced chemistry [124, 125]. This is key to the speculations, which we analyze in detail after the conclusions, that Titan might be capable of supporting some form of life [126, 127, 128, 129].

A fundamental question, then, is the stability of the vapor, liquid and solid phases of nitrogen-rich organic compounds in Titan's surface liquids. Each compound can enter the liquid either by leaving the solid phase or by leaving the atmosphere. Very little is known about the thermodynamic transitions of these molecules at such low temperatures. Data have been gathered using Fourier Transform Infrared Spectroscopy (FTIR) to analyze the composition of cryogenic methane solutions of acetonitrile and acetylene, but no conclusions have been drawn because of the difficulty of deconvoluting the FTIR spectra [130]. Some mixtures have been analyzed: the solubility of acetylene in methane at saturation has been measured down to 92 K [131, 132]. Vapor-liquid equilibria relevant to Titan's chemistry have been studied for the ternary system, methane-ethane-nitrogen, down to 95 K [133, 134, 135]. Depending on the surface ethane fraction, which is unclear on Titan, the fraction of dissolved nitrogen in the liquid at 95 K varies between 3.6% and 14.9% [135].

This lack of information is understandable: conducting experiments in cryogenic methane is not a simple task. This provides considerable incentive to employ our toolbox of theoretical approaches to determine the relevant thermodynamic properties. Approaches based on continuum thermodynamics have been used: The methane-ethane-propane-nitrogen-acetylene system has been studied using the van Laar model, which indicates that solid acetylene will be quite soluble under Titan surface conditions, with erosion similar to that of gypsum in water on Earth [136, 137]. The solubility of solid acetylene in pure liquid methane has been estimated using their pure-substance properties [138]. Similarly, the saturation solubility of HCN and acetylene in liquid ethane have been estimated using the enthalpy of fusion of pure HCN and acetylene [139]. The ternary methane-ethane-nitrogen system has been studied very recently us-

ing Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) [140]. And at the all-atom level, calculations of methane-nitrogen vapor-liquid equilibria at ultra-low temperatures have been previously performed using a theoretical approach known as Grand Canonical Monte Carlo (GCMC) simulation [141]. GCMC is a molecular-level simulation in which solute molecules are inserted into the solvent until an equilibrium is reached; here the chemical potential, volume and temperature are kept constant, but the number of particles fluctuates [141]. However, the computational efficiency of GCMC drops off rapidly with increasing solute size, making GCMC simulations difficult for some of the most interesting solutes on Titan.

In this chapter, we tackle the lack of information on the tendency of nitrogen compounds to dissolve in Titan's cold hydrocarbon seas. We investigate the solvation properties of the most common stable nitrogen compounds observed in Titan's upper atmosphere, and acetylene, using a full complement of techniques from equation of state approaches to all-atom Molecular Dynamics (MD). The equation of state approaches, performed by our colleagues Wael Fouad and Walter Chapman at Rice, include the Soave-Redlich-Kwong, Peng-Robinson, PC-SAFT, Polar PC-SAFT, and the Conductor-like Screening Model for Real Solvents (COSMO-RS). These methods are described in the next section.

6.1 Methods

Modeling the thermodynamics of nitrogen-containing systems at Titan is challenging for several reasons. First, nitriles exhibit strong polar interactions which can significantly influence their phase behavior, and most equations of state and

models do not explicitly take polar interactions into account. Second, the literature lacks any experimental data on the solubility of nitriles in liquid alkanes, either at Titan conditions or at higher temperatures and pressures. As a result, equations of state and other models are required to predict the needed solubilities must be used without the usual empirical fitting parameters.

Knowing how challenging the problem is, our colleagues Fouad and Chapman predicted the solubility of nitrogen-containing systems in pure liquid methane, methane-nitrogen and methane-ethane-nitrogen mixture using four different thermodynamic models. Peng-Robinson (PR) [142] and Soave-Redlich-Kwong (SRK) [143] equations of state have the advantage of being simple correlations of phase behavior that are fitted to reproduce the critical pressure and temperature of components under consideration. However, cubic equations of state do not predict accurate liquid phase densities, and they tend to perform poorly for associating and highly polar compounds, which is a severe problem for our polar solutes.

Another model used here is the conductor-like screening model for realistic solvents (COSMO-RS) [144, 145, 146], as implemented in the ADF molecular modeling suite [147, 148], which uses the COSMO approach developed by Klamt and Schuurmann [144] to predict activity coefficients through quantum calculations. The advantage of using COSMO-RS is that, without the need to fit parameters, it is as accurate as activity coefficient models fitted to large databases (*e.g.*, UNIFAC). The COSMO approach includes nearest neighbor polar interactions, but it cannot predict compressibility effects. This is probably acceptable on Titan's surface because the pressure is never far from the low pressure (by chemical industry standards) of 1.45 atm [149].

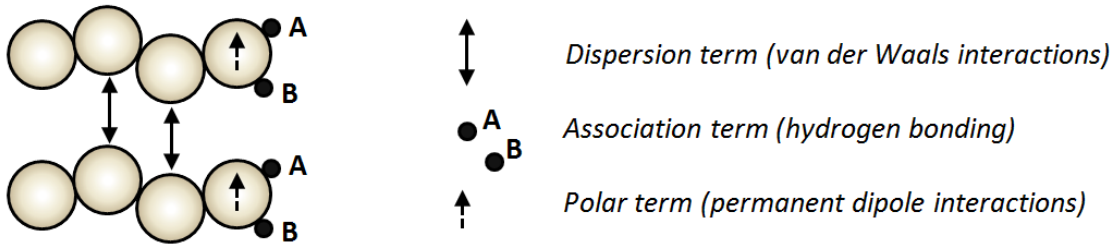


Figure 6.1: Schematic of the Polar and Perturbed Chain form of the Statistical Associating Fluid Theory (Polar PC-SAFT)

A final model to consider this the molecular equation of state developed by Chapman *et al.* [150, 151] for polar and associating fluids, known as the Statistical Associating Fluid Theory (SAFT). This theory includes association by using a coarse-grained atomic model based on Wertheim's first-order perturbation theory (TPT1) [152, 153, 154, 155, 156]. The polar and perturbed chain form of the equation of state (Polar PC-SAFT) [157, 158, 159], used here, models molecules as linear chains of polar hard spheres, as shown in Figure 6.1, with attached association sites that allow for hydrogen-bonding.

In Figure 6.2, we compare the performance of the models versus vapor-liquid equilibria data for some pure components of the Titan atmosphere (and also argon) at cryogenic temperatures. Figure 6.2 (a-b) shows that PC-SAFT was the most accurate of the models at predicting the shape of the phase diagram. The models performed equally well in predicting the liquid phase of the binary methane-nitrogen system in Figure 6.2 (c), while PC-SAFT showed superiority in predicting the composition of the vapor phase. COSMO-RS, as implemented by the ADF molecular modeling suite, does not allow molecular modeling of noble gas-containing systems; hence the model predictions were not included in Figure 6.2 (d-e). Although Figure 6.2b shows that COSMO-RS failed to predict

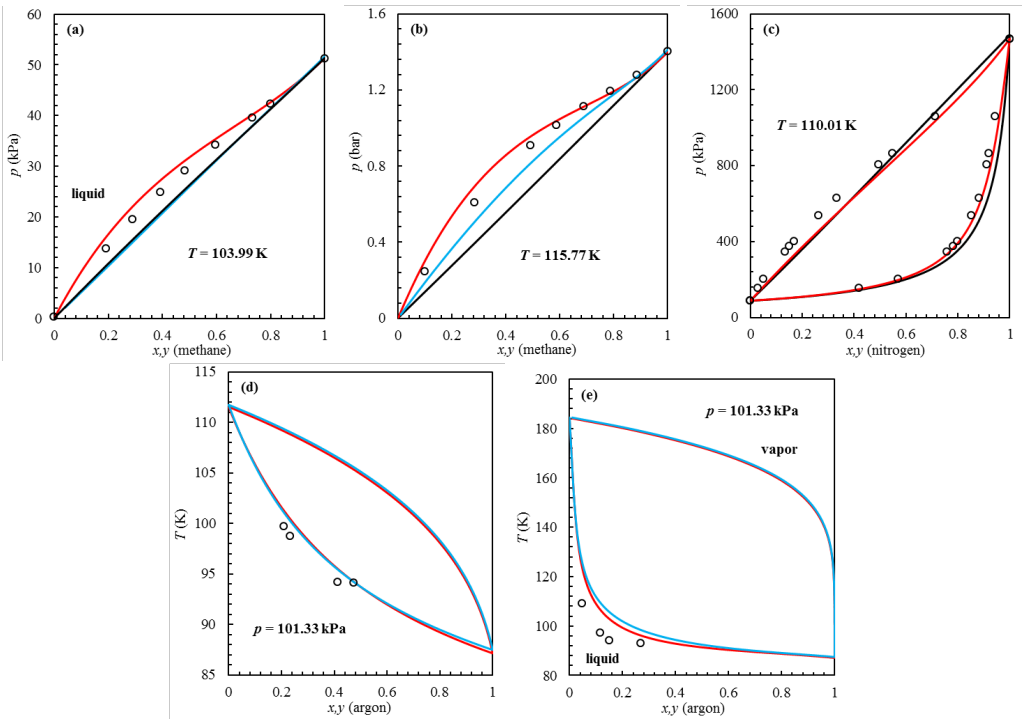


Figure 6.2: Vapor-liquid equilibrium (VLE) for light hydrocarbons and gases at conditions close to that on Titan. (a): methane-ethane [160]; (b): methane-propane [161]; (c): methane-nitrogen [162]; (d): methane-argon [163]; (e): ethane-argon [163]. (o): experimental data; red solid line (-): Polar PC-SAFT; black solid line (-): COSMO-RS; blue solid line (-): SRK.

the methane-propane interaction, this should not be taken as an indictment of its accuracy for the more complex systems we mean to examine. Van der Waals interactions are treated empirically and somewhat imprecisely within COSMO, which is mainly designed to compute short-range interactions through sigma profiles and electron density calculations, which in general do not adequately include dispersion [114]. As a result, we expect COSMO to perform less well for systems such as methane-propane where dispersion interactions dominate, achieving better results for polar compounds such as nitriles.

Comparison of the models with the small amount of related solubility data

which is available, mainly solid hydrocarbons in cryogenic liquids, is a necessary step in the validation process. The literature lacks any experimental data on the solubility of nitriles in liquid methane, so, for the sake of comparison, Dr. Fouad examined data found on the solubility of non-polar hydrocarbons, CO_2, and H_2S in liquid methane at 110 K. From a molecular perspective, methane can be expected to behave in a similar manner to the spherical molecules argon and krypton; for this reason, Dr. Fouad also compared the solubility of diethyl ether (moderately polar) in argon at 90 K as well as the solubility of acetone (quite polar) in krypton at 125 K. COSMO-RS cannot be used for noble gases, because it depends on generating the sigma profiles of the components, which is a delta function for noble gases. Therefore, unfortunately, the Ar and Kr and data cannot be compared to COSMO-RS calculations. The experimental solubilities are presented as ranges, as per Szczepaniec-Cieciak 2003 [164], in Table 6.2. As can be seen, all four models demonstrate an acceptable agreement with the experimental solubility range for solid non-polar hydrocarbons in liquid methane. On the other hand, models ignoring polar interactions (non-polar PC-SAFT, SRK, and PR) over-estimate the solubility of polar hydrocarbons in liquid argon and krypton. COSMO-RS results are not available for mixtures containing noble gases. The accuracy of Polar PC-SAFT for the solubilities of diethyl ether in argon and acetone in krypton shows that that the Polar PC-SAFT equation of state can be accurate for polar solutes in cryogenic nonpolar solvents.

Solute	Experiment	SRK	PR	PC-SAFT	COSMO-RS
n-Butane	10^{0}-10^{-1}	1.88×10^{-1}	2.08×10^{-1}	8.29×10^{-2}	3.97×10^{-1}
Cyclopentane	10^{-1}-10^{-2}	2.17×10^{-2}	2.87×10^{-2}	1.64×10^{-2}	7.75×10^{-1}
Hydrogen sulfide	10^{-2}-10^{-3}	1.83×10^{-3}	1.99×10^{-3}	1.18×10^{-3}	2.42×10^{-2}
n-Pentane	10^{-2}-10^{-3}	1.18×10^{-2}	1.36×10^{-2}	9.46×10^{-3}	1.27×10^{-1}
Carbon dioxide	10^{-3}-10^{-4}	4.31×10^{-4}	4.12×10^{-4}	4.28×10^{-4}	3.77×10^{-3}

Solute	Experiment	SRK	PR	PC-SAFT	COSMO-RS
n-Hexane	10^{-3}-10^{-4}	3.07×10^{-4}	3.32×10^{-4}	2.44×10^{-4}	5.14×10^{-3}
Cyclohexane	10^{-3}-10^{-4}	6.70×10^{-3}	6.17×10^{-3}	1.76×10^{-2}	1.75×10^{-1}
1,3-Butadiene	10^{-3}-10^{-4}	6.80×10^{-3}	7.52×10^{-3}	4.78×10^{-3}	2.64×10^{-2}
n-Heptane	10^{-4}-10^{-5}	1.04×10^{-4}	1.13×10^{-4}	6.57×10^{-5}	2.90×10^{-3}
Toluene	10^{-4}-10^{-5}	2.88×10^{-4}	4.00×10^{-4}	1.34×10^{-4}	4.09×10^{-2}
Diethyl ether *	10^{-5}-10^{-6}	2.13×10^{-3}	2.63×10^{-3}	1.31×10^{-4}	1.05×10^{-5}†
Acetone **	10^{-5}-10^{-6}	3.18×10^{-2}	4.10×10^{-2}	3.34×10^{-2}	6.57×10^{-5}†

Table 6.2: Solubility (mole fraction) of solid hydrocarbons in liquid methane at 110 K, liquid argon at 90 K and liquid krypton at 125 K. Experimental solubility ranges are taken from Szczepaniec-Cieciak [164]. *argon, **krypton. †Polar PC-SAFT, since COSMO-RS does not apply to noble gases.

As a final step in validating the pure-component SAFT parameters fitted in this work, Dr. Fouad examined the phase behavior of nitrile-containing systems at higher temperatures, where experimental data exist. Figure 6.3 demonstrates some of the phase diagrams of acetonitrile, acrylonitrile, propanenitrile and hydrogen cyanide-containing systems. It can be seen that cubic equations of state were unable to accurately predict the VLE in Figure 6.3 (a-c) even given a binary interaction parameter. In Figure 6.3a, Polar PC-SAFT and COSMO-RS performed similarly and were able to capture the non-ideality of the system. SRK predicted an unrealistic liquid-liquid phase split (not shown). Consequently, we concluded that the Polar PC-SAFT equation of state and the COSMO-RS model appear to be the best options in predicting the thermodynamics of nitrogen-containing systems for Titan applications.

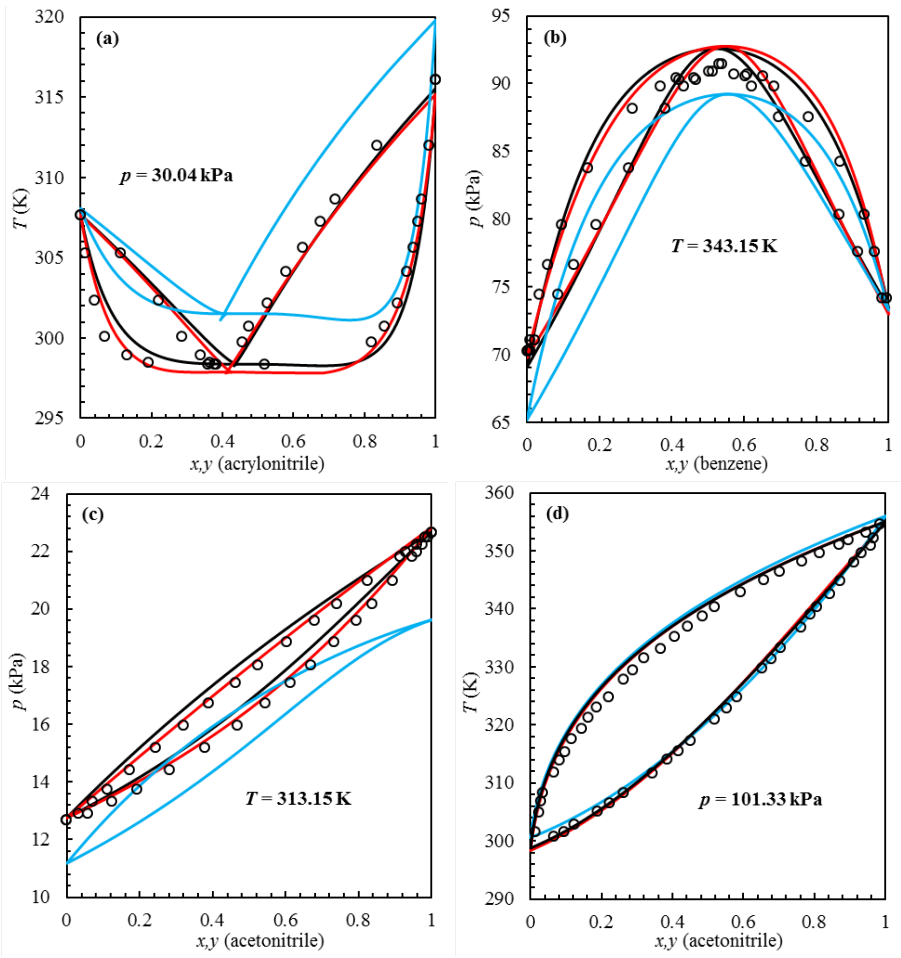


Figure 6.3: Vapor-liquid equilibrium (VLE) of nitriles containing systems at high temperatures. (a): acrylonitrile-n-hexane [165]; (b): acetonitrile-benzene [166]; (c): acetonitrile-propanenitrile [167]; (d): acetonitrile-hydrogen cyanide [168]. (o): experimental data; red solid line (-): Polar PC-SAFT; black solid line (-): COSMO-RS; blue solid line (-): SRK.

Based on the validation process, the solubility of nitriles and acetylene in pure liquid methane at Titan conditions can be predicted with some confidence using the thermodynamic models described here. The binary interaction parameters k_{ij} between the solutes and solvent had to be set to zero since experimental data to parameterize them is not available (the effect of this choice is described later in Tables 6.8 and 6.9). Dr. Fouad extended the models to predict

the solubilities in 10 mole percent nitrogen in liquid methane (mixture 1) as well as in a methane-ethane-nitrogen ternary mixture (mixture 2) on the high end of expected ethane concentrations: 50.1% methane, 41.3% ethane, 8.6% nitrogen [136]. This makes mixture 2 a test of whether ethane concentration will have a significant effect on the solubility of polar nitrogen compounds. We used a temperature of 94 K, close to typical polar Titan surface temperatures of 92 K and within measured seasonal variation [123]. It is possible that the solubility in this system will show an unusually strong dependence on temperature, because the temperatures are very close to the triple point of methane, 91 K, where the solid phase of methane begins play a role. However, we note that liquid-phase properties tend to vary only slowly across the triple point in supercooled liquids, and that the presence of dissolved nitrogen in Titan's surface liquids will tend to reduce the freezing point. The *critical* point, by contrast, has a strong effect on liquid-phase properties when it is crossed, but at 191 K the critical temperature is far outside Titan's temperature range.

We also calculated the standard Gibbs free energies of solvation, ΔG°_{solv}, for the transition from vacuum to solvent for both liquid methane and nitrogen gas solvents using Molecular Dynamics. We calculated ΔG°_{solv} using two well-established methods, Free Energy Perturbation (FEP) [169, 170, 171] and Thermodynamic Integration (TI) [172, 173, 174]. These free energy methods require a mathematical description of the interactions between each of the atoms in the system, often termed a force field; we used the Optimized Potential for Liquid Simulations, OPLS, a force field well-proven for use in TI and FEP calculations [174, 170, 171, 175]. OPLS is a particularly suitable choice since it was optimized to describe the behavior of organic molecules, especially in hydrocarbon solvents like those found on Titan [85]. We used the standard OPLS parame-

ters unchanged from the Jorgensen *et al.* 2008 data sets [176], in the all-atom form which is most accurate for free-energy calculations [174]. For the nitrogen atmosphere, we used the Lennard-Jones parameters of Bouanich for nitrogen gas at 100 K [177]. We validated the effectiveness of OPLS for our study by confirming the OPLS-generated structures and pairwise binding energies of all the molecules we considered against *ab initio* quantum mechanical calculations [178]. We excluded any molecules which produced errors in bond distances of more than 0.2 Angstrom, in bond angles, 1°, in dihedral angles, 2°, and in binding energies, 1 kcal/mol.

6.1.1 Fugacity of the solid phase

In the equation-of-state calculations performed by our colleagues Dr. Fouad and Professor Chapman at Rice, accurate determination of the fugacity of the solid phase plays a significant role. The relationship between the fugacities of the solid and solution phases can be expressed as:

$$\ln \frac{f_i^S}{f_i^{L0}} = \frac{H_i^f}{RT_i^f}\left(1 - \frac{T_i^f}{T}\right) \tag{6.1}$$

where f_i^S and f_i^{L0} are the fugacities of the solute in the solid and the sub-cooled liquid phase, and H_i^f and T_i^f are the heat of fusion and the melting temperature of the solute, respectively.

The fugacity of a component can be represented as a function of temperature and pressure:

$$f_i^S\left(T,p\right)=f_i^S\left(T,p^{\text{sat}}\right)\exp\int_{p^{\text{sat}}}^{p}\frac{v_i^S}{\text{RT}}\text{dp} \tag{6.2}$$

and

$$f_i^{L0}\left(T,p\right)=f_i^{L0}\left(T,p^{\text{sat}}\right)\exp\int_{p^{\text{sat}}}^{p}\frac{v_i^{L0}}{\text{RT}}\text{dp} \tag{6.3}$$

Taking the ratio of equations 6.2 and 6.3 leads to

$$\frac{f_i^S}{f_i^{L0}}\left(T,p\right)=\frac{f_i^S}{f_i^{L0}}\left(T,p^{\text{sat}}\right)\exp\int_{p^{\text{sat}}}^{p}\frac{v_i^S-v_i^{L0}}{\text{RT}}\text{dp} \tag{6.4}$$

Inserting equation 6.1 into equation 6.4 leads to the following expression for the fugacity of the solute in the solid phase:

$$f_i^S\left(T,p\right)=f_i^{L0}\left(T,p\right)\exp\left[\int_{p^{\text{sat}}}^{p}\frac{v_i^S-v_i^{L0}}{\text{RT}}\text{dp}+\frac{H_i^f}{RT_i^f}\left(1-\frac{T_i^f}{T}\right)\right] \tag{6.5}$$

6.1.2 Polar PC-SAFT Equation of State

The complete equation of state in Polar PC-SAFT, in terms of the Helmholtz free energy, is the sum of an ideal gas contribution, a^{id}, a hard-chain contribution, a^{hc}, a dispersion contribution, a^{disp}, an association contribution, a^{assoc}, and a polar contribution a^{polar}:

$$a^{res}\equiv a-a^{id}=a^{hc}+a^{disp}+a^{assoc}+a^{polar} \tag{6.6}$$

For the hard chain contribution, Chapman *et al.* [150, 151] developed an equation of state applicable for homonuclear hard-sphere chains comprising m segments, given by

$$a^{hc} = \dot{m}a^{hs} - \sum_i x_i \left(m_i - 1\right) \rho \ln\left(g_{ii}^{hs} \sigma_{ii}\right) \tag{6.7}$$

and

$$\dot{m} = \sum_i x_i m_i \tag{6.8}$$

where x_i is the mole fraction of chains of component i, m_i is the number of segments in a chain of component i, ρ is the number density of molecules, g_{ii}^{hs} is the radial pair distribution function for segments of component i in the hard sphere system, and the superscripts hc and hs indicate quantities of the hard-chain and hard-sphere systems, respectively. For mixtures of hard spheres [179, 180]:

$$a^{hs} = \frac{1}{\xi_0}\left[\frac{3\xi_1\xi_2}{1-\xi_3} + \frac{\xi_2^3}{\xi_3\left(1-\xi_3\right)^2} + \left(\frac{\xi_2^3}{\xi_3^2} - \xi_0\right)\ln\left(1-\xi_3\right)\right] \tag{6.9}$$

and

$$g_{ii}^{hs} = \frac{1}{1-\xi_3} + \left(\frac{d_i d_j}{d_i + d_j}\right)\frac{3\xi_2}{1-\xi_3} + \left(\frac{d_i d_j}{d_i + d_j}\right)^2 \frac{3\xi_2^2}{\left(1-\xi_3\right)^3} \tag{6.10}$$

where

$$\xi_n = \frac{\pi}{6}\rho \sum_i x_i m_i d_i^n \tag{6.11}$$

Based on Barker and Henderson's perturbation theory [181], the temperature-dependent segment diameter $d_i(T)$ is

$$d_i(T) = \sigma_i \left[1 - 0.12 \exp\left(\frac{-3\epsilon_i}{kT}\right)\right] \tag{6.12}$$

where ϵ_i is the depth of square-well potential and k is the Boltzmann constant.

The perturbation theory of Barker and Henderson is used to calculate the attractive part of the chain interactions. It is a second-order theory where the dispersion term of the Helmholtz free energy, A^{disp}, is given as a sum of first- and second- order contributions

$$a^{disp} = \frac{A_1}{NkT} + \frac{A_2}{NkT} \tag{6.13}$$

where N is Avogadro's number. Gross and Sadowski [182, 183] derived a set of equations appropriate for a square-well potential function

$$a^{disp} = -2\pi\rho \left[I_1(\eta, \dot{m})\right] \left(m^2 \epsilon \sigma^3\right)_m - \pi\rho\dot{m} C_1 \left[I_2(\eta, \dot{m})\right] \left(m^2 \epsilon^2 \sigma^3\right)_m \tag{6.14}$$

and

$$C_1 = \left(1 + Z^{hc} + \rho \frac{\partial Z^{hc}}{\partial \rho}\right)^{-1} \tag{6.15}$$

The value of C_1 represents the compressibility of the hard chain fluid, $I_1(\eta, \dot{m})$ and $I_2(\eta, \dot{m})$ are power series in density where the coefficients are functions of the chain length and the packing fraction, η, which is equivalent to ξ_3:

$$I_1(\eta, \dot{m}) = \sum_{j=0}^{6} a_j(\dot{m})\eta^j \tag{6.16}$$

$$I_2(\eta, \dot{m}) = \sum_{j=0}^{6} b_j(\dot{m})\eta^j \tag{6.17}$$

The parameters of a pair of unlike segments, σ_{ij} and ϵ_{ij}, are obtained using standard Berthelot-Lorentz mixing rules.

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \tag{6.18}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}\left(1 - k_{ij}\right) \tag{6.19}$$

In Eq. 6.13, the mixing rules for $\left(m^2\epsilon\sigma^3\right)_m$ and $\left(m^2\epsilon^2\sigma^3\right)_m$ are defined by

$$\left(m^2\epsilon\sigma^3\right)_m = \sum_{i}^{nc}\sum_{j}^{nc} x_i x_j m_i m_j \left(\frac{\epsilon_{ij}}{kT}\right)\sigma_{ij}^3 \tag{6.20}$$

$$\left(m^2\epsilon^2\sigma^3\right)_m = \sum_{i}^{nc}\sum_{j}^{nc} x_i x_j m_i m_j \left(\frac{\epsilon_{ij}}{kT}\right)^2 \sigma_{ij}^3 \tag{6.21}$$

The components modeled in this work were all assumed to be non-associating, but we include a description of the association term to explain the options avail-

able. Chapman *et al.* [150, 151] derived the association term Helmholtz free energy, A^{assoc}, for mixtures:

$$a^{assoc} = \sum_i X_i \left[\sum_{A_i} \left[\ln X^{A_i} - \frac{X^{A_i}}{2} \right] + \frac{1}{2} M_i \right] \tag{6.22}$$

where M_i is the number of the association sites per molecule and the mole fraction of molecules that are not bonded at site A, X^{A_i}, can be determined as follows

$$X^{A_i} = \left(1 + \sum_j \sum_{B_j} \rho_i X^{B_j} \Delta^{A_i B_j} \right)^{-1} \tag{6.23}$$

where $\Delta^{A_i B_j}$ is the "association strength" defined as

$$\Delta^{A_i B_j} = d_{ij}^3 g_{ij}(d_{ij})^{seg} \kappa^{A_i B_j} \left[\exp\left(\frac{\epsilon^{A_i B_j}}{kT} \right) - 1 \right] \tag{6.24}$$

The mixing rules suggested by Wolbach and Sandler [184] are used for the association energy, $\epsilon^{A_i B_j}$, and the effective association volume, $\kappa^{A_i B_j}$:

$$\epsilon^{A_i B_j} = \frac{1}{2} \left(\epsilon^{A_i B_i} + \epsilon^{A_j B_j} \right) \tag{6.25}$$

$$\kappa^{A_i B_j} = \sqrt{\kappa^{A_i B_i} \kappa^{A_j B_j}} \left(\frac{\sqrt{\sigma_{ii} \sigma_{jj}}}{0.5(\sigma_{ii} + \sigma_{jj})} \right)^3 \tag{6.26}$$

The change in free energy due to polar interactions, as derived by Jog and Chapman [158, 159], is accurately obtained by dissolving all the bonds in a chain

and then applying the u-expansion to the resulting mixture of polar and nonpolar spherical segments. The polar contribution written in the Padé approximate has the following form:

$$a^{polar} = \frac{a_2}{1 - a_3/a_2} \tag{6.27}$$

where a_2 and a_3 are the second-order and third-order terms in the perturbation expansion. Written for mixtures, and allowing for multiple dipolar segments, these terms have the following form:

$$a_2 = \frac{-2\pi\rho}{9(kT)^2} \sum_i \sum_j x_i x_j m_i m_j x_{p_i} x_{p_j} \frac{\mu_i^2 \mu_j^2}{d_{ij}^2} I_{2,ij} \tag{6.28}$$

$$a_3 = \frac{5\pi\rho^2}{162(kT)^3} \sum_i \sum_j x_i x_j m_i m_j x_{p_i} x_{p_j} \frac{\mu_i^2 \mu_j^2 \mu_k^2}{d_{ij}^2 d_{jk}^2 d_{ik}^2} I_{3,ijk} \tag{6.29}$$

In the aforementioned equations, $I_{2,ij}$ and $I_{3,ijk}$ are the angular pair and triplet correlation functions and μ_i is the dipole moment for component i. As shown by Jog *et al.* [158, 159], they are related to the corresponding pure fluid integrals by

$$I_{2,ij} = I_2(\eta, \dot{m}) \tag{6.30}$$

$$I_{3,ijk} = I_3(\eta, \dot{m}) \tag{6.31}$$

6.1.3 Cubic Equations of State

In 1972, Soave proposed an improved, generalized version of the Redlick-Kwong equation (RK) [185] by introducing Pitzer's acentric factor (ω) into the attraction term. Soave redefined the $\alpha(T_R, \omega)$ function while keeping the RK volume functionality as it is. The following summarizes the general equations found in what is now called the Soave-Redlick-Kwong (SRK) [143] equation of state.

$$P = \frac{RT}{V-b} - \frac{a_c \alpha(T_R, \omega)}{V(V+b)} \tag{6.32}$$

$$a_c = 0.42747 \frac{RT_c^{2.5}}{P_c} \tag{6.33}$$

$$b = 0.08664 \frac{RT_c}{P_c} \tag{6.34}$$

$$\alpha(T_R, \omega) = \left[1 + m\left(1 - T_R^{(}0.5)\right)\right]^2 \tag{6.35}$$

$$m = 0.480 + 1.574\omega - 0.176\omega^2 \tag{6.36}$$

In 1976, Peng and Robinson improved upon Soave's equation by recalculating the $\alpha(T_R, \omega)$ function and by modifying the volume-dependence of the attraction term. These changes resulted in better predictions of liquid volumes and vapor-liquid equilibrium (VLE) for non-polar hydrocarbons. The Peng-Robinson equation of state (PR) [142] is summarized as the following:

$$P = \frac{RT}{V-b} - \frac{a_c \alpha\left(T_R, \omega\right)}{V(V+b)+b(V-b)} \tag{6.37}$$

$$a_c = 0.45724 \frac{RT_c^{2.5}}{P_c} \tag{6.38}$$

$$b = 0.07780 \frac{RT_c}{P_c} \tag{6.39}$$

$$\alpha\left(T_R, \omega\right) = \left[1 + m\left(1 - T_R^{(}0.5)\right)\right]^2 \tag{6.40}$$

$$m = 0.37464 + 1.54226\omega - 0.26992\omega^2 \tag{6.41}$$

In both the traditional versions of SRK and PR, van der Waals one (vdW-1) mixing rules are applied to multi-component systems

$$a = \sum_i \sum_j x_i x_j a_{ij} \tag{6.42}$$

$$b = \sum_i \sum_j x_i x_j b_{ij} \tag{6.43}$$

$$a_{ij} = \sqrt{a_i a_j}\left(1 - k_{ij}\right) \tag{6.44}$$

6.1.4 Parameterizations of the Equations of State

Based on the work done by Cordier *et al.* [186], Tan *et al.* [187] and Glein and Shock [188], Titan lakes are estimated to contain mixtures of methane, ethane, propane, nitrogen and traces of acetylene. As part of the validation process, Dr. Fouad compared how the models perform in predicting the vapor-liquid equilibrium (VLE) of binary mixtures containing light hydrocarbons and inert gases at conditions close to that on Titan. Binary interaction parameters, k_{ij}, were set to zero in this validation step. Pure component critical properties and acentric factors used within the SRK and PR framework can be found in Table 6.3. Antoine equation parameters required for pure component vapor pressure calculation needed for the COSMO approach are illustrated in Table 6.4. In addition, pure component parameters fit to saturated liquid density and vapor pressure data [189] and used in the SAFT framework can be found in Table 6.5 for the non-polar parameters and 6.6 for the parameters specific to Polar PC-SAFT. Our binary interaction parameters, fitted to experimental data, are shown in Table 6.7. Data to set the binary interaction parameters are not available for all species, and thus these were set to the default value of 0.0. An analysis of the effect of changing these parameters is shown in Table 6.8 and Table 6.9.

Component	T_c (K)	P_c (bar)	ω	T_m (K)	ΔH_{fus} (kJ mol^{-1})
Methane	190.58	46.04	0.011	90.67	0.94
1,3-Butadiene	425.37	43.30	0.193	164.25	7.984
Acetone	508.40	47.80	0.304	177.80	5.687
Acetonitrile	545.50	48.33	0.338	229.32	8.16
Acetylene	308.32	61.39	0.187	192.40	3.76
Acrylonitrile	535.00	44.80	0.350	189.67	6.23

Component	T_c (K)	P_c (bar)	ω	T_m (K)	ΔH_{fus} (kJ mol^{-1})
Argon	150.86	48.98	0.000	83.80	1.18
Benzene	562.16	48.98	0.211	278.68	9.87
Carbon dioxide	304.19	73.82	0.228	216.58	8.652
Cyclohexane	553.54	40.75	0.212	279.69	2.677
Cyclopentane	511.80	45.20	0.196	179.35	0.607
Diethyl ether	466.70	36.38	0.285	156.85	7.190
Ethane	305.32	48.72	0.099	90.36	2.72
Hydrogen cyanide	456.65	53.91	0.410	259.91	8.41
Hydrogen sulfide	373.53	89.63	0.083	187.68	2.375
Krypton	209.35	55.02	0.000	115.78	1.640
n-Butane	425.18	37.97	0.199	134.86	4.664
n-Heptane	540.26	27.36	0.351	182.57	14.037
n-Hexane	507.43	30.12	0.305	177.84	13.08
Nitrogen	126.10	33.94	0.040	63.15	0.36
n-Pentane	469.65	33.69	0.249	143.42	8.401
Propane	369.83	42.48	0.152	85.52	3.5
Propanenitrile	564.40	41.80	0.324	180.40	5.03
Toluene	591.79	41.09	0.264	178.18	6.636

Table 6.3: Critical and melting properties of components used in this work [190, 191]

Component	A	B	C	
Methane	3.9895	443.028	-0.49	[192]
Acetonitrile	4.27873	1355.37	-37.853	[193]
Acrylonitrile	4.06661	1255.94	-41.853	[194]

Component	A	B	C	
Benzene	4.01814	1203.83	-53.226	[195]
Ethane	4.50706	791.3	-6.422	[196]
Hydrogen cyanide	4.67417	1340.79	-11.592	[197]
n-Hexane	4.00266	1171.53	-48.784	[195]
Nitrogen	3.7362	264.651	-6.788	[198]
Propane	4.01158	834.26	-22.763	[199]
Propanenitrile	3.61732	1036.42	-83.76	[200]

Table 6.4: Antoine equation parameters for components used in this work

Component	m	σ (Angstrom)	ϵ/k (K)	Ref
Methane	1	3.7039	150.03	[182]
n-Hexane	3.0576	3.7983	236.77	[182]
Nitrogen	1.2053	3.313	90.96	[182]
Argon	0.9285	3.4784	122.23	[182]
Benzene	2.4653	3.6478	287.35	[182]
Acetylene	2.157	2.906	168.551	[187]
n-Butane	2.3316	3.7086	222.88	[182]
Cyclopentane	2.3655	3.7114	265.83	[182]
Hydrogen sulfide	1.6941	3.0214	226.79	[201]
n-Pentane	2.6896	3.7729	231.2	[182]
Carbon dioxide	2.0729	2.7852	169.21	[182]
n-Hexane	3.0576	3.7983	236.77	[182]
Cyclohexane	2.5303	3.8499	278.11	[182]
1,3-Butadiene	2.2309	3.5892	228.6	[201]

Component	m	σ (Angstrom)	ϵ/k (K)	Ref
n-Heptane	3.4831	3.8049	238.4	[182]
Toluene	2.8149	3.7169	285.69	[182]
Krypton	0.961	3.6634	167.26	[201]
Diethyl ether	2.9686	3.5147	220.09	[22, 24]
Acetone	2.1873	3.6028	245.49	[201, 202]
Acetonitrile	1.911	3.371	351.26	This work
Propanenitrile	2.229	3.532	328.04	This work
Acrylonitrile	1.999	3.544	332.46	This work
Hydrogen cyanide	1.372	3.404	368.9	This work

Table 6.5: Pure component parameters for the non-polar terms in the PC-SAFT equation of state

Component	m	σ (Angstrom)	ϵ/k (K)	x_p	μ (D)	Ref
Diethyl ether	2.8787	3.5549	220.59	0.3474	1.2	[22, 24]
Acetone	2.221	3.6079	259.99	0.2258	2.7	[201, 202]
Acetonitrile	1.0703	4.1857	392.85	0.4217	3.92	This work
Propanenitrile	1.3868	4.2022	290.34	0.4821	4.04	This work
Acrylonitrile	1.272	4.219	326.884	0.626	3.236	This work
Hydrogen cyanide	1.0673	3.7654	305.5	0.5721	2.98	This work

Table 6.6: Pure component parameters specific to the Polar PC-SAFT equation of state

System	k_{ij}, SRK	k_{ij}, Polar PC-SAFT
Acrylonitrile + n-Hexane	0.114	-0.0313

System	k_{ij}, SRK	k_{ij}, Polar PC-SAFT
Acetonitrile + Benzene	0.0745	-0.058
Acetonitrile + propanenitrile	0.0189	-0.0156
Acetonitrile + Hydrogen cyanide	-0.048	-0.05

Table 6.7: Binary interaction parameters used in the SRK and Polar PC-SAFT framework

Table 6.8 and Table 6.9 show an analysis of the effect of changing k_{ij} on the calculated solubilities from PC-SAFT. The sensitivity, unfortunately, is high, indicating that the unknown values of k_{ij} for these interactions (which would normally be parameterized *via* experimental data) is a possible avenue for error in the calculations from PC-SAFT. Most likely the true k_{ij} values are slightly below our value of 0.0. This would make the calculated solubilities somewhat larger. The trends in solubility by species should remain the same.

System / k_{ij}	-0.2	-0.1	0.0	0.1	0.2
Acetonitrile	1.2543×10^{-3}	6.0493×10^{-6}	3.0413×10^{-8}	1.5767×10^{-10}	7.9992×10^{-13}
Acrylonitrile	0.1654	1.5388×10^{-3}	4.6570×10^{-6}	1.4848×10^{-8}	4.8967×10^{-11}
Propanonitrile	0.4390	8.8149×10^{-3}	1.3266×10^{-5}	2.3568×10^{-8}	4.3365×10^{-11}
Acetylene	0.4029	0.1439	5.2824×10^{-3}	1.8777×10^{-4}	7.0120×10^{-6}

Table 6.8: Solubility in liquid methane using non-polar PC-SAFT (94 K, 1.5 bar), as a function of the binary interaction parameter k_{ij}

System / k_{ij}	-0.2	-0.1	0.0	0.1	0.2
Acetonitrile	1.0511×10^{-6}	1.3063×10^{-8}	1.6676×10^{-10}	2.1536×10^{-12}	5.8176×10^{-14}
Acrylonitrile	8.4793×10^{-8}	1.1336×10^{-9}	1.5541×10^{-11}	1.7453×10^{-13}	5.8176×10^{-14}
Propanenitrile	2.8883×10^{-7}	2.2883×10^{-9}	1.8568×10^{-11}	1.7453×10^{-13}	5.8176×10^{-14}

Table 6.9: Solubility in liquid methane using polar PC-SAFT (94 K, 1.5 bar), as a function of the binary interaction parameter k_{ij}

6.1.5 Molecular Dynamics Simulations

Free Energy Perturbation (FEP) [169, 170, 171] and Thermodynamic Integration (TI) [172, 173, 174] are based on a thermodynamic path between the vacuum and the interacting state for each test molecule, with an interaction parameter, λ, varying from 0 to 1 to indicate the degree of interaction. In TI, we calculate the derivative of the potential energy, $dE/d\lambda$, at every step along a pre-determined path, then average over all microstates to provide $dG/d\lambda$ as a function of λ. Once we have established values for $dG/d\lambda$, it is strightforward to integrate numerically over λ to give ΔG°_{solv} [174]. TI is not a thermodynamically sophisticated method, but it has consistently been found to be accurate in calculating free energy changes [174].

In FEP, each small change in λ is regarded as a change in the sampling of the same set of available states; the free energy change of the system between two values, λ_1 and λ_2, is then given by the Zwanzig equation:

$$\Delta G = -k_B T \ln\left(< \exp\left(\frac{E(\lambda_1) - E(\lambda_2)}{k_B T} \right) > \right) \tag{6.45}$$

In which <angle brackets> indicate an ensemble thermodynamic average. FEP is formally exact for any change in λ, but works well in practice only when the difference is a small perturbation, because its accuracy declines relatively quickly when the microstates sampled at λ_1 do not overlap with the microstates at λ_2 [203]. Because of this deficiency, FEP has not been found in practice to be as accurate as TI. However, FEP has the useful property that it can be separately performed over the same path both forward and in reverse [170]. This means that sampling errors can be identified. Thus, we use FEP as a means of confirming, and providing error bounds to, our TI results.

We performed tests of the OPLS parameters using *ab initio* Density Functional Theory (DFT) calculations at the computationally exacting, but particularly accurate, M062X/aug-cc-pVDZ level of theory [204]. This "Minnesota functional" is known to yield accurate pairwise binding energies for van der Waals and polar systems [205]. We performed the *ab initio* quantum mechanical calculations in Gaussian 09 using the Berny optimization algorithm, [206] with the M062X/aug-cc/pVDZ functional in the presence of an implicit hydrocarbon solvent [207] and in vacuum.

OPLS models for cyanoacetylene and 2,4-pentadiynenitrile proved to be inaccurate in terms of the reproduction of their bond lengths, and for cyanoallene in their C-H bond lengths. This is probably due to the highly unsaturated carbon atoms in these molecules, which also make these molecules less stable and less likely to persist outside the upper atmosphere. Thus, they were rejected from further study. For the remaining species, the average difference

Species	Δ(Bond length) (nm)	Δ(Bond angle) (°)	Δ(Dihedral) (°)
Propanenitrile	0.0004	0.5	0.2
Methane	0.0022	1.5	0.0
HCN	0.0005	0.0	0.0
Cyanoacetylene	0.0047	0.0	0.0
Acrylonitrile	0.0008	1.3	0.0
Cyanoallene	0.0007	0.9	0.0
Acetonitrile	0.0004	0.1	0.0
HC_5N	0.0059	0.0	0.0

Table 6.10: Error in OPLS-optimized structures. Average absolute difference, for each species, between its OPLS structure and its structure *via ab initio* optimization with implicit solvent.

between the OPLS and *ab initio*-generated structures for the bond lengths was 0.01 Angstrom, in the angles, 0.6°, and in the dihedrals (where present) 0.7°. A list of all these properties for the molecules we tested is provided in Table 6.10.

Once the veracity of OPLS-generated structures was confirmed against *ab initio* simulations, the binding energies were also validated in a similar manner. Two sets of *ab initio* binding energies were created for comparison, one in vacuum and one in implicit solvent. The solvent-based energies are closer to experimental conditions, while the energies in vacuum were used essentially as error bounds. The difference between these energies was between 0.1 and 1.3 kcal/mol for all species under evaluation, with an average of 0.6 kcal/mol. The difference between the OPLS and *ab initio* binding energies ranged from 0.2 to 0.9 kcal/mol, again averaging 0.6 kcal/mol. In general, values for the OPLS-generated binding energy for a given species fell between the *ab initio* values with and without solvent. These results validate use of the OPLS models; the error associated with using the OPLS force field is smaller than the fluctuation due to the presence of solvent. An outlier among the molecules we studied was HCN, which had an OPLS binding energy that was 1.8 kcal/mol above the

Species	OPLS (kcal/mol)	DFT_{vacuum} (kcal/mol)	$DFT_{solvent}$ (kcal/mol)
Propanenitrile	-6.4	-7.5	-6.2
Methane	0.0	0.0	0.0
HCN	-4.1	-3.0	-2.3
Acrylonitrile	-5.6	-6.0	-4.9
Acetonitrile	-5.9	-6.3	-5.0

Table 6.11: Pairwise binding energies. OPLS and *ab initio* calculations produced similar results.

value for the binding energy predicted by an *ab initio* representation with an implicit solvent, suggesting that the OPLS model of HCN is somewhat inaccurate. Binding energies found by the three methods are provided in Table 6.11.

For each molecule tested, we used three simulation "boxes" to represent the three different phases of interest: a 100 Angstrom cube containing 95% N_2 and 5% methane gas representative of the Titan atmosphere, a 20 Angstrom cube of liquid methane, and a 6x6x6 molecule box of the pure solid test molecule, sized based on the size of the test molecule. We find that allowing a full relaxation of the crystal solid is difficult and that, in order to simulate the solid correctly, it would have be initialized already close to the true crystal geometry. Doing otherwise would lead to higher-energy, amorphous solid phases, which would cause an underestimation of the $\Delta G_{solv,solid}$.

The test molecule's interaction with its surroundings, considering both Coulombic and Lennard-Jones forces, was linearly reduced to zero by multiplication by the scaling factor λ over the course of 100 steps. This is not the most efficient path - reducing Coulombic and Lennard-Jones interactions separately would have required fewer steps - but this path made the free energy calculations very straightforward. Each of the 100 steps was a separate simulation of 0.5 ns, for a total simulation time of 50 ns per species. The sampling error was checked by

the comparison between TI and FEP to ensure that sufficient sampling had been performed.

All Molecular Dynamics runs were performed using the Large-scale Atomic / Molecular Massively Parallel Simulator (LAMMPS) [95]. A constant-temperature, constant-pressure (NPT) ensemble was applied using the Nose-Hoover thermostat and barostat [208]. The temperature was set to 94 K and the pressure to 1.45 atm, conditions found on the surface of Titan [149, 209]. A Verlet integration scheme was used with a time step of 2 fs [208]. The Particle-Particle Particle-Mesh algorithm was used to add long-range Coulombic interactions, which are significant in membrane simulations [210]. The short-range Coulombic interactions and Lennard-Jones interactions were computed pairwise with a cut-off radius of 8 Angstrom [210].

6.2 Solvation Results

As can be seen in Table 6.12, solubilities predicted by our colleagues Fouad and Chapman using COSMO-RS, SRK and PR for the pure methane case agree reasonably with each other, with the exception of hydrogen cyanide. In reference to Table 6.12, changing the liquid phase composition had only a minor effect on the solubility of the nitriles and acetylene in mixture 1. In terms of mixture 2, both SRK and PR equations of state predicted higher nitriles solubility in comparison to the pure methane case. On the other hand, COSMO-RS showed a slight decrease in the solubility. In their interactions with polar compounds, such as the solubility of water in supercritical methane and ethane, these two solvents are known to behave very similarly [211]. Therefore, it is unsurprising that the

solubility of nitriles is only moderately affected by the addition of ethane.

	Methane	Mix. 1	Mix. 2
SRK (1.5 bar)			
Acetylene	1.64×10^{-2}	1.03×10^{-2}	2.87×10^{-2} **
Acetonitrile	5.18×10^{-6}	2.51×10^{-6}	3.65×10^{-5}
Propanenitrile	4.76×10^{-4}	2.31×10^{-4}	4.02×10^{-3}
Acrylonitrile	7.53×10^{-6}	4.82×10^{-5}	8.00×10^{-4} **
Hydrogen Cyanide	1.78×10^{-6}	1.08×10^{-6}	1.38×10^{-5} **
PR (1.5 bar)			
Acetylene	1.82×10^{-2}	1.16×10^{-2}	2.99×10^{-2} **
Acetonitrile	8.91×10^{-6}	4.63×10^{-6}	5.29×10^{-5}
Propanenitrile	7.71×10^{-4}	4.04×10^{-4}	5.44×10^{-3}
Acrylonitrile	1.27×10^{-5}	8.49×10^{-5}	1.10×10^{-3}
Hydrogen Cyanide	2.92×10^{-6}	1.85×10^{-6}	1.92×10^{-5}
COSMO-RS (1.01325 bar)			
Acetylene	1.57×10^{-2}	1.58×10^{-2}	9.11×10^{-4}
Acetonitrile	7.46×10^{-7}	6.36×10^{-7}	2.18×10^{-7}
Propanenitrile	5.05×10^{-4}	4.16×10^{-4}	1.42×10^{-4}
Acrylonitrile	2.76×10^{-5}	2.42×10^{-5}	9.57×10^{-6}
Hydrogen Cyanide	7.30×10^{-8}	7.58×10^{-8}	3.58×10^{-8}

Table 6.12: Solubility (mole fraction) of nitriles and acetylene in pure methane, methane-nitrogen and methane-ethane-nitrogen mixtures using SRK, PR and COSMO-RS at 94 K. Mix. 1 = 90% methane, 10% nitrogen by mole. Mix. 2 = 50.1% methane, 41.3% ethane, 8.6% nitrogen by mole. **Indicates three-phase (vapor-liquid-solid) equilibrium.

In terms of the SAFT framework, solubilities for nitrogen-containing compounds and acetylene are reported in Table 6.13 using the polar and non-polar versions of PC-SAFT. Based on the results in Table 6.12, we do not expect the solubilities to be dramatically different in nitrogen or ethane mixtures than in pure methane. As previously shown in Table 6.2, adding the polar term to the perturbation theory led to a dramatic decrease in the solubilities in comparison to predictions found in Table 6.12. This is reasonable because the solutes are polar and the solvent non-polar, so adding the polar correction to PC-SAFT emphasizes their incompatibility. Based on the previous argument, we would expect the Polar PC-SAFT equation of state to predict the most reliable solubilities in comparison to the other models used. However, in the absence of experimental data, it is challenging to make a definitive conclusion. As shown in Tables 6.8 and 6.9, the PC-SAFT model is somewhat sensitive to the value of the binary interaction parameter k_{ij}, but the experimental data needed to calculate k_{ij} values is not available. This problem makes us consider COSMO-RS as arguably the most accurate of the available models that we considered, because it does not rely on the parameterization of k_{ij}.

Component	Polar PC-SAFT	PC-SAFT
Acetylene	5.32×10^{-3}	5.32×10^{-3}
Acetonitrile	1.67×10^{-10}	3.04×10^{-8}
Propanenitrile	1.86×10^{-11}	1.33×10^{-5}
Acrylonitrile	1.55×10^{-11}	4.66×10^{-6}
Hydrogen cyanide	5.67×10^{-12}	1.12×10^{-11}

Table 6.13: Solubility (mole fraction) of nitriles and acetylene in pure methane at 94 K and 1.5 bar, with and without including the polar term in the PC-SAFT framework.

The results of the free energy calculations for ΔG°_{solv} are provided in Table 6.14. The value of ΔG°_{solv} of the atmosphere is low, but non-zero, and represents the deviation of Titan's atmosphere from an ideal gas, which can be seen to be minimal. ΔG°_{solv} for the atmosphere increases with increasing molecule size, as expected.

Evaluating the differences between values of ΔG°_{solv} in Table 6.14 provides the free ΔG° necessary to pass between phases. The degree of preference of each test molecule for the liquid over the gas is roughly proportional to its molecular weight. All of the values are significant and negative, meaning that each of these substances prefers being in the liquid state rather than form part of the atmosphere. This is consistent with the fact that Titan's atmosphere becomes supersaturated with nitrogen compounds and forms aerosols of condensed liquids [149, 209]. Unfortunately, the property ΔG°_{solv} solid, which would allow the calculation of solubility directly from the free energy calculations, is not cur-

ΔG°_{solv} atmosphere:	$FEP_{\rightarrow}$ (kcal/mol)	TI (kcal/mol)	$FEP_{\leftarrow}$ (kcal/mol)
Propanenitrile	-0.035	-0.035	-0.033
Acrylonitrile	-0.043	-0.043	-0.042
Acetonitrile	-0.025	-0.025	-0.024
Acetylene	-0.016	-0.016	-0.016
HCN	-0.017	-0.017	-0.016
ΔG°_{solv} sea:	$FEP_{\rightarrow}$ (kcal/mol)	TI (kcal/mol)	$FEP_{\leftarrow}$ (kcal/mol)
Propanenitrile	-3.17	-3.19	-3.21
Acrylonitrile	-5.89	-5.67	-4.74
Acetonitrile	-3.93	-3.74	-3.15
Acetylene	-2.59	-2.44	-1.78
HCN	-2.84	-2.68	-2.23

Table 6.14: Free energies of solvation with respect to different thermodynamic states. The TI column should be considered most accurate, with the FEP columns interpretable as bounds for the sampling error.

rently accessible: the crystal states of the solutes must be successfully simulated first. Our simulations of the solid phase showed amorphous structures and densities which were systematically too low by more than 10%, indicating a failure to find the true global minimum structures. Therefore, the calculation of Titan's solubility properties from all-atom MD simulation must be reserved for future work when crystal structures are available.

Validating models for nitrogen-containing compounds in methane and its mixtures, and then calculating liquid and vapor solubilities, is a considerable challenge in the absence of experimental data. Our approach was to use the most accurate theoretical models that are available and, at least, determine bounds on the most likely values of these properties. Not surprisingly, the agreement between the different solubility calculations is imperfect. For example, COSMO-RS and Polar PC-SAFT, two of the most theoretically advanced models available, vary by a factor of almost three in predicting the solubility of acetylene in pure methane, with COSMO-RS producing the higher value. COSMO-RS is

similar to PR and SRK in this prediction, which may indicate that it shares a weakness with the simpler cubic equations of state. For polar interactions, we have evidence from experimental data in argon and krypton that use of Polar PC-SAFT should be accurate for polar compounds to within an order of magnitude or better. On the other hand, PC-SAFT models rely on the binary interaction parameter k_{ij}, which for many compounds of interest is not available. This uncertainty reinforces the need for experimental data on solubility in cryogenic systems.

The MD calculations did not produce solubility data comparable to those of the other models because of the difficulty of simulating the correct crystal structures of the solid phases. Comparison between the stability of the solid phase and the dissolved phase is possible in MD simulation; but it can only be carried out if the crystalline simulation is initialized with the correct experimental crystal structure. Predicting the correct crystal structure is a significant challenge by itself, and we found that attempting to create the crystal from an amorphous phase was not effective. However, the MD simulations did provide important insight into the system, because the free energies they provided can be related to the known structure of the molecules and simple considerations of charge and Van der Waals forces. Calculation of ΔG°_{solv} for the solid using MD simulation would make the sources of error in the other means of calculation more clear, and is a useful subject for further study.

As to the composition of Titan's seas, and its impact on solubility, we can derive some confidence from the fact that the solubilities calculated by COSMO-RS are not sensitive to the presence of dissolved nitrogen. For acetylene, say, the difference between pure methane and methane+N_2 is less than 1%. A large

concentration of ethane, however (41%), has a stronger impact; for acetylene, as much as an order of magnitude. Therefore, the ethane concentration in Titan's seas can be seen as the most significant unknown fact about their composition from a solubility perspective.

The solubilities of the nitrogen compounds, as predicted by COSMO-RS and Polar PC-SAFT, are extremely small, showing that nitrogen groups are extremely effective at reducing solubility in the Titan environment. Non-polar ethane, for example, is capable of being dissolved in Titan's seas at concentrations of over 10%, while the same molecule with the substitution of a polar nitrogen group (acetonitrile) is over one million times less soluble. This difference is much greater than that between soluble and insoluble minerals in Earth systems; rock salt and quartz differ in solubility by a factor of less than one million [212, 213]. This supports our hypothesis that insoluble nitrogen compounds will play a special and important role in Titan's geochemistry.

6.3 Azotosomes: Life in nonpolar solvents

The lipid bilayer membrane which is the foundation of life on earth is not viable outside of biology based on liquid water. This fact has caused astronomers who seek conditions suitable for life to search for exoplanets within the "habitable zone," the narrow band in which liquid water can exist. But can cell membranes be created and function at temperatures far below those at which water is a liquid? We take a step toward answering this question by proposing a new type of membrane, composed of small organic nitrogen compounds, which is capable of forming and functioning in liquid methane at cryogenic tempera-

tures. Using molecular simulations, we demonstrate that these membranes in cryogenic solvent have an elasticity equal to that of lipid bilayers in water at room temperature. As a proof-of-concept, we also demonstrate that stable cryogenic membranes could arise from compounds observed in the atmosphere of Saturn's moon, Titan, known for the existence of seas of liquid methane on its surface.

Studies of proto-biology and the formation of cells on earth generally focus on nucleic acids; indeed, it is generally believed that ribonucleic acids were the precursors to terrestrial life (the "RNA world" hypothesis) [214]. However, recent research has shown that RNA catalysis depends on establishing high local concentrations, meaning that compartmentalization was likely a co-equal or earlier phase than RNA [215]. This supports a "lipid world" hypothesis, in which lipid membranes played an important role as a very early evolutionary step in creating life on Earth [216]. Lipid bilayers, without additional cellular machinery, have been observed to grow, divide, aid polymerization reactions, and even synthesize RNA from polymerase enzymes [217, 218, 219].

Terrestrial cell membranes are composed of a bilayer of phospholipids; surfactants composed of nonpolar lipid chains and oxygen-laden polar heads. The polar heads form surfaces compatible with water, allowing the membrane to separate the aqueous world outside and the aqueous life within. The lipid tails of the phospholipids aggregate by van der Waals forces, thus stabilizing the membrane. A vesicle made from such a membrane is known as a liposome.

The role of self-assembled surfactants in evolutionary biology on Earth raises the question of whether non-aqueous conditions can support any analogous structure. Experimental studies have been performed to create vesicles in non-

polar solvents. This has included the consideration of non-ionic ethers [220], esters [221], surfactants [222] and inverted phospholipids [223]. Inverse phospholipid membranes have even been considered as biological possibilities in liquid methane [224, 225, 226, 227, 228, 220, 221, 222, 223, 229].

However, the phospholipid membranes which are so strong and elastic in water do not perform as well in liquid methane. The non-polar tails of phospholipids would seem to be compatible with nonpolar liquid methane, and the polar heads with each other; does this suggest that an inverse membrane to those formed in water could exist in methane? Unfortunately not; this phospholipid hypothesis neglects the fact that the tails of phospholipids are long-chain hydrocarbons, which will be rigid at cryogenic temperatures. Furthermore, the phospholipid head component atoms, oxygen and phosphorous, are not available in any form in the methane seas of Titan, and presumably not in any similar liquid methane environment. Inverted phase liposomes are therefore not a viable option. The idea of using polarity to prevent dissolution, however, is valid if any suitable materials exist.

As a proof-of-concept for vesicle formation in a methane-rich environment, we began our search for polar materials using those which form naturally in Titan's methane-nitrogen atmosphere, as shown in Table 6.15 [225, p. 167]. Lower in the atmosphere, all of these species condense into aerosols, preventing further observation by Cassini, but laboratory experiments to reproduce the methane-nitrogen atmosphere have generally produced a tar-like residue of molecules, called tholins [230]. These tholins have been found to consist of hydrocarbons, nitriles, and amines [230]. Therefore, we have also included primary nitriles and amines of lengths propyl-hexyl in our study, although the abundance of the

tholins relative to the Cassini-observed species is uncertain.

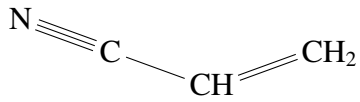
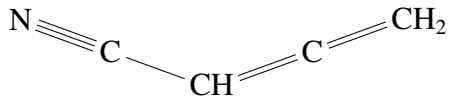
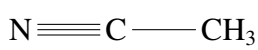
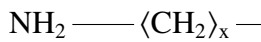
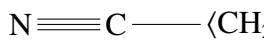
Name	Structure	C_i (ppm)
HCN		200
Cyanoacetylene		40
Acrylonitrile		10
Cyanoallene		4
Acetonitrile		3
2,4-pentadiynenitrile		1
Propanenitrile		0.5
Primary amine		0
Primary nitrile		0

Table 6.15: Polar nitrogen compounds found on Titan and their abundance in the upper atmosphere as measured by Cassini.[225, p. 167]

Given the challenges of experimental studies at cryogenic temperatures, we once again adopted a molecular simulation approach to screen for the most promising candidates for self-assembly into a structure reminiscent of a membrane. We considered only short ligands, given the fact that longer ligands offer no advantage at such cold temperatures. Our candidate molecules are all much shorter than typical phospholipids, which involve carbon chains 15-20 atoms long. Liquid methane is cold enough to solidify almost any substance: a 4-membered carbon chain, butane for example, is far below its freezing point of 133 K in liquid methane. At such temperatures, it might seem almost impossible for a flexible organic membrane to form, let alone one with similar flexibility to that of a lipid bilayer. However, low temperature also allows small molecules

to aggregate differently than they would at 300 K. Our normal intuition needs to be adjusted for a colder world.

We hypothesized that liquid methane membranes would rely on the polarity of nitrogen-containing groups ("azoto-" groups) to hold them together, in the same way that terrestrial liposomes rely on the non-polarity of alkyl groups. Therefore, we referred to these structures as "azotosomes." A comparison between the liposome and proposed azotosome structures is shown in Figure 6.4.

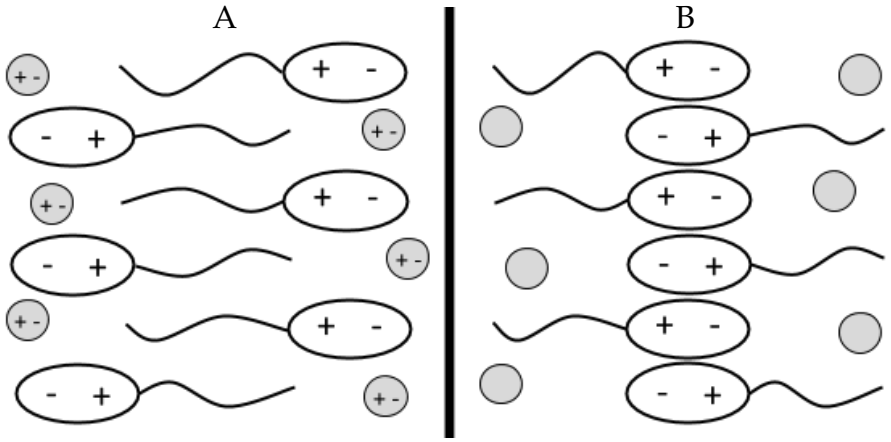


Figure 6.4: Liposomes and azotosomes **A)** liposome in polar solvent. Polar heads are braced by non-polar lipid tails. **B)** azotosome in non-polar solvent. Non-polar tails are braced by polar nitrogen-rich heads.

The key physical requirements of a membrane are that it be flexible and stable. The most common measure of cell membrane flexibility is the area expansion modulus K_a (also known as the dilational, stretch, or area compressibility modulus) [231, 232]. The area expansion modulus of terrestrial cell membranes at room temperature is 0.24-0.50 J/m^2 [231, 232]. As we will show, several of our azotosome candidates lie within this range. The most common measure of stability is the decomposition energy or stability timescale [233]. In general, lipid bilayers on Earth are metastable [234]. We will show that some of our azo-

tosome candidates have high decomposition energies relative to the cryogenic environment, resulting in very long stability timescales.

Synthesizing azotosomes for experimental study would be a challenging project, akin to the first synthesis of liposomes and with the added difficulty of cryogenic conditions. However, the molecules that compose azotosomes are similar to molecules routinely studied on earth, making the properties of azotosomes accessible through standard molecular simulation. Molecular Dynamics (MD) has been used for decades to simulate bilayer membranes [235, 236, 237, 233] and polymer vesicles (polymersomes) [238]. The values of K_a calculated by MD have been found to agree closely with experiments, [231] supporting our use of the technique for these new membranes.

To provide the molecular forces in our simulations we used Optimized Potentials for Liquid Simulations (OPLS), which are well known and effective models for liquid hydrocarbons, small organic molecules, [239, 178] and polymersomes [238]. We validated OPLS models for our molecules by confirming the OPLS-generated structures and pairwise binding energies against quantum mechanical calculations, [178] as described in Materials and Methods. Only the molecules that OPLS modeled faithfully in terms of bond lengths, angles, and binding energies were passed on for study as azotosome candidates.

The first property of azotosomes we investigated was their flexibility, as measured by the area expansion modulus K_a. One procedure that has been used to calculate this property is simulation of the spontaneous fluctuation of a membrane's area over a lengthy MD run [231]. However, this method relies on random sampling of rare events, and thus has problems with accuracy, overestimating K_a by a factor of two or more [231]. Instead, we indented the

membrane by a given amount and calculated the area expansion modulus from the resistance to stretching [231]. This is directly comparable to the nanoindentation method which is used to find K_a experimentally, and comparison of such simulations and experiments on terrestrial liposomes has shown them to yield equivalent results [231].

The second criterion that we used to assess azotosomes was their stability. One procedure that has been used to study liposome stability is to simulate the membrane over a lengthy MD run and observe whether it dissociates spontaneously [233]. Waiting for dissociation is, unfortunately, only effective for the weakest membranes, since MD typically covers only nanoseconds of simulated time. A superior procedure is to find the timescale of dissociation *via* the activation energy [240]. By Arrhenius's Law, the rate of a process with activation energy ΔE is proportional to exp(-ΔE / k_BT), where k_B is the Boltzmann constant and T is the absolute temperature [240]. If the time scale of the process without an energy barrier, t_0, is known, then the time scale of the process with the barrier is given by t = t_0 exp(ΔE / k_BT).

The dependence of the time scale on ΔE is exponential, so the uncertainty in the time scale will be greater than the underlying uncertainty in the energy barrier. Fortunately, we only require the time scale's order of magnitude. For a molecule to move one Angstrom at 94 K with no energy barrier takes on the order of t_0 = 1 picosecond. If an activation energy of 8 kcal/mol is required, for instance, this gives a timescale of 1 picosecond $\times$ exp(8 kcal/mol/k_BT) = 235,000 seconds (three Earth days). In the absence of model error, this time scale is sufficiently long that it could lead us to suggest that an azotosome is sufficiently stable if it has a decomposition energy barrier higher than 8 kcal/mol. With a 20%

uncertainty, corresponding to the error between the OPLS and *ab initio* binding energies of the azotosome components, a more conservative cutoff would be to use a barrier of 10 kcal/mol, corresponding to a decomposition time of over 100 Earth years.

In order to find the flexibility and energy barriers of our candidate azotosomes, we need to calculate all possible routes by which a molecule can leave its azotosome and then sum the relevant properties of these routes according to the probability of their occurrence. The properties we chose as relevant here are the potential energy and the force on the test molecule in the direction of the membrane. To prepare all these possible routes, we began with a representative section of membrane, a 6×6 x-y grid of molecules, for each candidate species.

This piece of membrane was then expanded periodically to simulate a 2D membrane of arbitrary extent. The surrounding space was filled with methane solvent. The edge molecules (those for which x=0 or y=0) were held fixed in the z-direction to keep the membrane in place. The membrane was allowed to equilibrate, self-assembling into its desired structure as shown in Figure 6.5.

Next, a test molecule (at the location, x, y = 3,3) was incrementally withdrawn in the z-direction. At each increment, the test molecule was allowed to move freely in the [x, y] directions. In the z-direction, it was loosely restrained by a harmonic potential, allowing it to sample nearby z-locations, but not to leave the vicinity. This procedure, known as "umbrella sampling," allows a sampling of all possible configurations, and also provides a measure of whether the sampling was adequate - we can simply check whether we have enough samples from each small increment in the z-direction. A schematic of this process is shown in Figure 6.6.

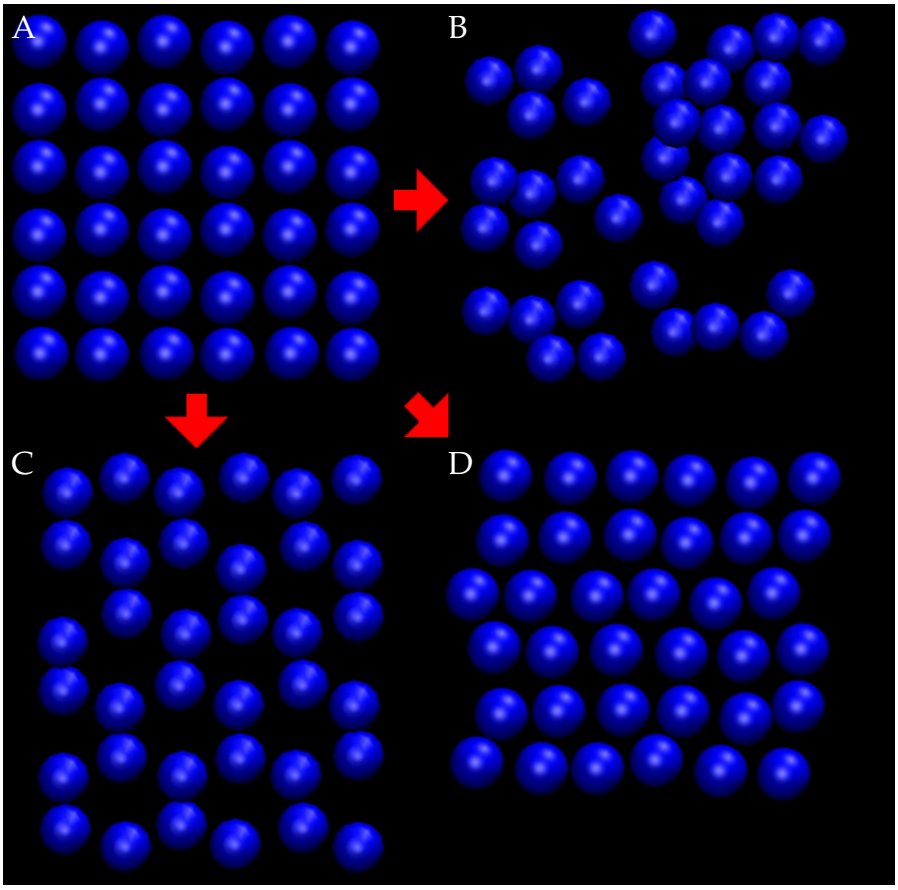


Figure 6.5: Nitrogen-head positions in selected azotosomes. **A)** The initial grid, **B)** aminopentane (amorphous), **C)** pentanenitrile (hexagonal), **D)** acrylonitrile (close packed hexagonal).

This procedure is necessary because of the azotosomes' high barriers to decomposition. If the barriers to decomposition were low, the system would naturally sample the decomposed states within the simulated time. Umbrella sampling forces the simulation to reach these states even if they would require a very long time to arise from inherent fluctuations.

At each of the millions of molecular dynamics steps, the potential energy function was evaluated, and the force required to hold the test molecule away from the membrane was calculated. This yielded profiles of force and potential energy as a function of distance from the membrane, which are provided

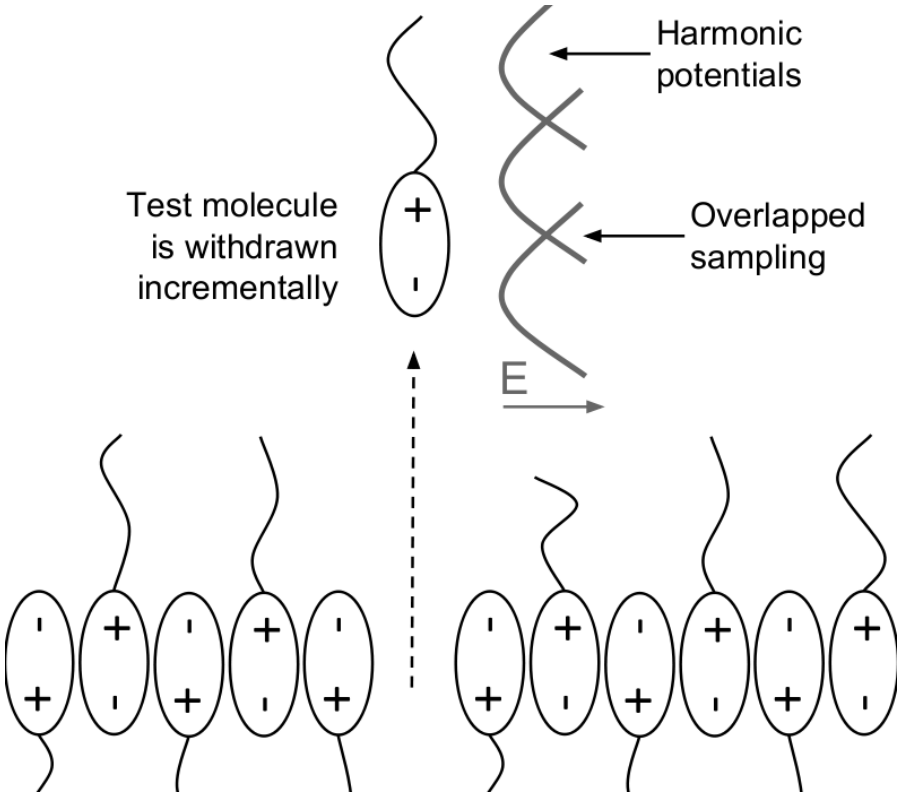


Figure 6.6: Umbrella sampling of the azotosome decomposition process. The test molecule is incrementally withdrawn from the membrane in the z-direction.

separately in the Supplemental Information.

The force profiles were integrated over distance to yield free energy profiles, as shown in Figure 6.7. The change in area was calculated by measuring the deflection of the indented nitrogen atom at the center of the sheet and finding the area of the rectangular pyramid formed between it and the restrained nitrogen atom at the corners of the sheet. This procedure proved to be very robust with respect to changes in the overall sheet size. The deflection of the indented atom was measured relative to the equilibrium state of the membrane; that is, relative to the point of minimum free energy.

The potential energy barrier for each azotosome was calculated by finding

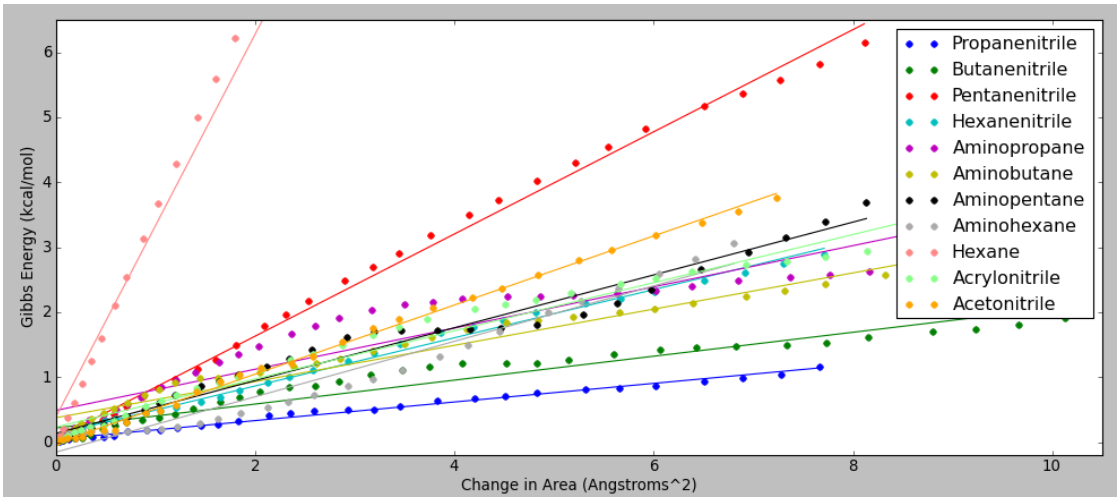


Figure 6.7: Stretching the azotosomes. The slope of the fit line is proportional to the area modulus K_a.

the largest single uninterrupted increase in the potential energy during the decomposition of each azotosome. The concept of a rate-determining, Arrhenius-style energy barrier relies on the barrier being uninterrupted. If there is a stable intermediate state at which the system can re-equilibrate, then it is not one barrier, but two smaller barriers, with a drastic increase in the implied rate of reaction. Since we used umbrella sampling, with its fine-grained view of the energy profiles, we were able to divide them into very fine sections (0.05 Angstrom) and make sure that there were no intermediate states within our barriers. An example barrier (acrylonitrile) is shown in Figure 6.8.

A few of the fine 0.05 Angstrom sections, almost all at the beginning and end where no umbrella sampling overlap could occur, had fewer than 500 samples. We removed these sections for consideration as barriers since (500 samples)×(2 femtoseconds/sample) is smaller than the timescale of our NPT equilibration (1000 femtoseconds). This has the effect of making our energy barrier calculations more conservative, since the sampling frequency of a section is inversely

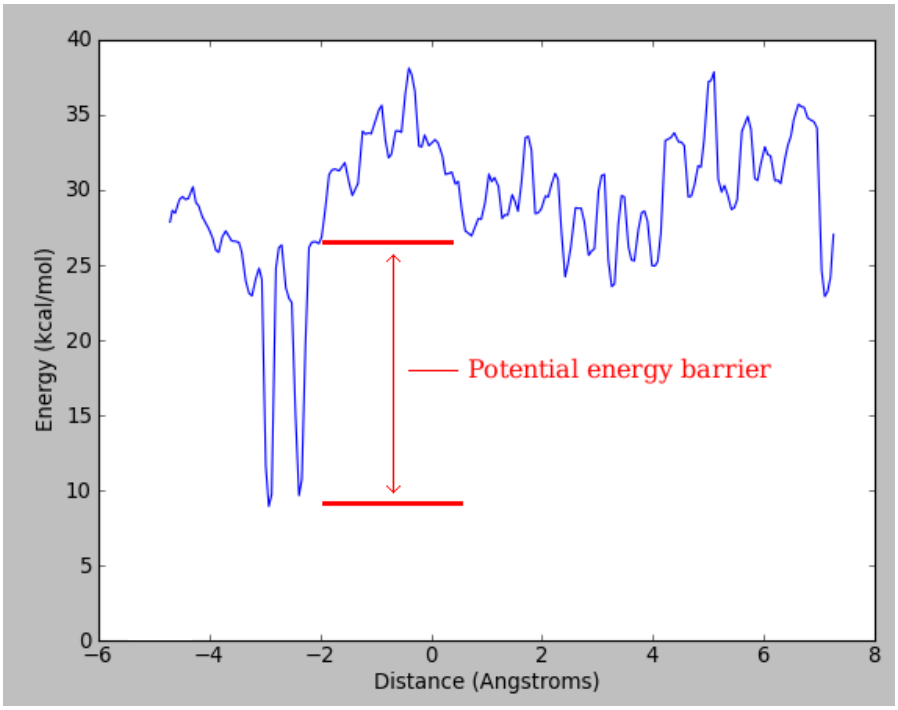


Figure 6.8: Potential energy profile for the decomposition of acrylonitrile. The largest instantaneous energy barrier is the activation energy to decompose the azotosome.

related to its energy.

All Molecular Dynamics runs were performed using the Large-scale Atomic / Molecular Massively Parallel Simulator (LAMMPS), distribution March 2014 [95]. Standard OPLS parameters were used unchanged from the Jorgensen *et al.* 2008 data sets [241]. Typical best practices for the MD simulation of membranes were followed. For example, a constant-temperature, constant-pressure (NPT) ensemble was applied using the Nose-Hoover thermostat and barostat; the barostat being anisotropic [242]. The anisotropic barostat allows the membrane the freedom to change or lose its structure, increasing the realism of the simulation [233]. The temperature was set to 94 K and the pressure to 1.45 atm, similar to the conditions on the surface of Titan [230]. Verlet integration was

used with a time step of 2 fs [242]. The Particle-Particle Particle-Mesh algorithm was used to add long-range Coulombic interactions, which are significant in membrane simulations [243]. The short-range forces, Coulombic interactions, and Lennard-Jones forces were computed pairwise with a cut-off radius of 8 Angstrom [243]. Each membrane was initialized as a flat x-y plane with periodic boundary conditions in x, y, and z, as shown in Figure 6.5. The initial simulation cell size was 21×21×42 Angstrom. The position of the test molecule was altered in increments of 0.2 Angstrom, starting at 2 Angstrom below its starting position in the membrane and ending 10 Angstrom above, giving 60 simulations per species tested. Data collection was then performed for 1 ns at each increment. The z-directional restraint on each molecule was exerted on its terminal nitrogen atom or, in a non-polar molecule, on its terminal carbon. This allowed each molecule to rotate if it experienced an asymmetric force, giving it access to the maximum number of degrees of freedom. The restraints on the edge molecules were absolute (z-forces were set to zero), while the restraint on the test molecule was imposed as a harmonic force with a spring constant of 10 kcal/mol-Angstrom. This moderate restraint allowed the test molecule to cover the 0.2 Angstrom range and overlap with the ranges of other simulations, providing continuous sampling of the properties of the system.

The errors in OPLS models for the bond lengths, bond angles, and dihedral (torsion) angles for the compounds we considered as azotosome candidates are provided in Table 6.16.

Calculations for the three most highly unsaturated compounds - cyanoacetylene, cyanoallene, and 2,4-pentadiynenitrile - were not considered further, because their OPLS models exhibited errors in the average bond lengths of over

Species	Δ(Bond length) (nm)	Δ(Bond angle) (°)	Δ(Dihedral) (°)
Propanenitrile	0.000	0.5	0.0
Butanenitrile	0.001	0.5	0.1
Pentanenitrile	0.001	0.4	0.1
Hexanenitrile	0.000	0.6	0.2
Aminopropane	0.001	0.7	0.7
Aminobutane	0.001	0.6	0.6
Aminopentane	0.001	0.6	0.4
Aminohexane	0.001	0.5	0.2
Hexane	0.001	0.3	0.0
HCN	0.001	0.0	0.0
Acrylonitrile	0.001	1.2	0.0
Acetonitrile	0.000	0.0	0.0
Cyanoacetylene	0.005	6.3	0.0
Cyanoallene	0.008	9.5	0.0
2,4-pentadiynenitrile	0.006	75.8	0.0

Table 6.16: Error in OPLS-optimized structures. Average difference, for each species, between its OPLS structure and the structure obtained *via ab initio* optimization with Onsager's self-consistent reaction field (SCRF) model of an implicit solvent.

0.04 Angstrom, which we considered to be unacceptably large. Being highly unsaturated, these species are the least likely to be stable when they leave the upper atmosphere anyway, so their unsuitability for OPLS modeling is unlikely to be important here. For the remaining compounds, the average difference between the OPLS and *ab initio* structures for the bond lengths was 0.008 Angstrom, 0.6° in the angles and 0.3° in the dihedrals (where present).

Ab initio binding energies were generated in vacuum and in implicit solvent. The difference between these energies was between 0.1 and 1.3 kcal/mol for all species under evaluation with an average of 0.6 kcal/mol. The difference between the OPLS and *ab initio* binding energies ranged from 0.2 to 0.9 kcal/mol, averaging 0.6 kcal/mol. In general, values for the OPLS-generated binding energy for a given species fell between the *ab initio* values with and without sol-

vent, which supports the use of the OPLS models. OPLS pairwise binding energies would be expected to fall between those of *ab initio* in vacuum and *ab initio* in solvent because the OPLS pairwise calculations are "solvent-like" due to the non-polarizability of the OPLS model. This makes the OPLS model experience a partial solvent effect, even when only a pair of molecules is present. As we noted earlier, HCN binding energies are simulated poorly by OPLS. However, as we shall show later, HCN does not self-assemble into an azotosome due to its small size. Its excessive binding energy will thus be irrelevant since it will not be considered further due to its lack of self-assembly. Binding energies found by the three methods (OPLS and *ab initio* calculations for the system in vacuum and solvent) are provided in Table 6.11.

Each azotosome begins the simulation as a grid of molecules, then self-assembles into its preferred structure. Species which OPLS could not represent accurately were removed from consideration, as was HCN, which did not form an ordered layer, and hexane, which formed a solid. The area expansion moduli of the remaining azotosomes, and also a plain hexane bilayer with no functional head, are shown in Table 6.17. Given the accuracy of the OPLS binding energies, we expect these values of K_a and ΔE to be accurate to within 20%.

All of our azotosomes have flexibilities similar to those of known terrestrial cell membranes: 0.13-0.55 J/m^2 for the azotosomes vs. 0.24-0.50 J/m^2 for terrestrial liposomes. With respect to thermal fluctuations, azotosomes will appear stiffer than Earth liposomes, because the thermal fluctuations on Titan are smaller than those on Earth. But with respect to mechanical stress, cryogenic azotosomes and room-temperature liposomes will respond surprisingly alike.

The comparison between hexane and hexanenitrile is instructive. Plain hex-

Species	K_a (J/m^2)	ΔE (kcal/mol)
Acrylonitrile	0.26	17.1
Acetonitrile	0.37	5.8
Propanenitrile	0.10	7.6
Butanenitrile	0.13	6.4
Pentanenitrile	0.55	8.4
Hexanenitrile	0.26	6.0
Aminopropane	0.22	5.2
Aminobutane	0.19	6.7
Aminopentane	0.28	7.7
Aminohexane	0.30	7.7
Hexane bilayer	2.20	—

Table 6.17: Flexibility K_a of nitrile and amine azotosomes, and activation energy ΔE to remove a molecule from each azotosome.

ane forms a layer eight times stiffer than hexanenitrile. Furthermore, the hexane bilayer is brittle, as shown in Figure 6.9. After a small amount of stretching, it appears to snap. In contrast, the hexanenitrile layer stretches smoothly throughout. The only difference between these two compounds is that hexanenitrile has a polar nitrogen head.

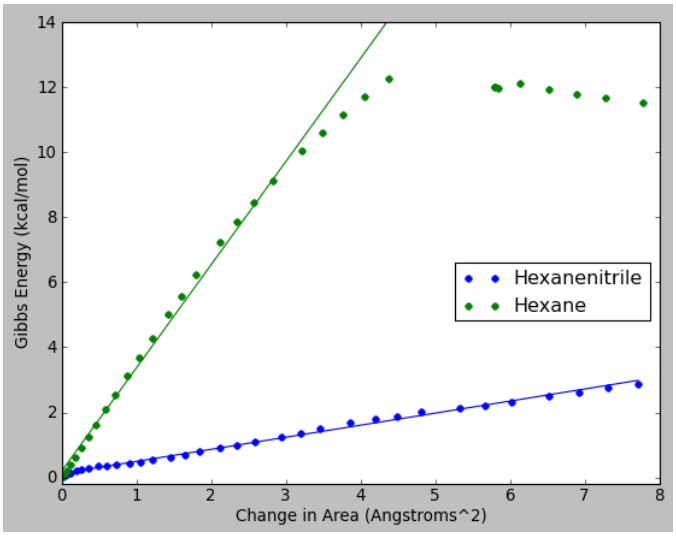


Figure 6.9: Stretching a hexanenitrile azotosome and a hexane bilayer. The slope of the linear fit is proportional to the area modulus K_a.

The key difference between the pure hydrocarbon layer and an azotosome is the structure resulting from the polar nitrogen head. We believe that it is this structure that allows a cryogenic azotosome to have the flexibility of a room-temperature lipid bilayer.

The dissociation energy barriers for each azotosome are also given in Table 6.17. Acetonitrile, butanenitrile, hexanenitrile, aminopropane, and aminobutane have ΔE values well below 8 kcal/mol, indicating unstable azotosomes. Propanenitrile, pentanenitrile, aminopentane, and aminohexane azotosomes have energy barriers close to 8 kcal/mol. They should be considered possible candidates, because their values lie within a 20% uncertainty that we feel is appropriate here. Acrylonitrile azotosomes show high barriers to decomposition, 17 kcal/mol, that are sufficient to ensure their stability over long time scales.

The geometry of the acrylonitrile molecule appears to favor the azotosome and hamper other states (Figure 6.10). This aligns with the fact that acrylonitrile is known experimentally to have a somewhat disordered solid phase [244]. Like all the azotosomes studied here, acrylonitrile azotosomes are symmetrical with respect to the plane of the membrane, implying that -like a lipid bilayer- they should be able to form vesicles of many sizes.

The calculation of free energies, required to determine K_a, also allowed us to derive the total free energy of decomposition for each azotosome. These values are shown in Table 6.18. All of these free energies are positive, indicating that the azotosome state is preferable to the dissolved state. These free energy values are concentration-dependent, since $\Delta G_{dissolve}$ is always negative at infinite dilution. The concentration created by dissolving one azotosome molecule in each box is about 0.1%, which is higher than we might expect in Titan's seas. However,

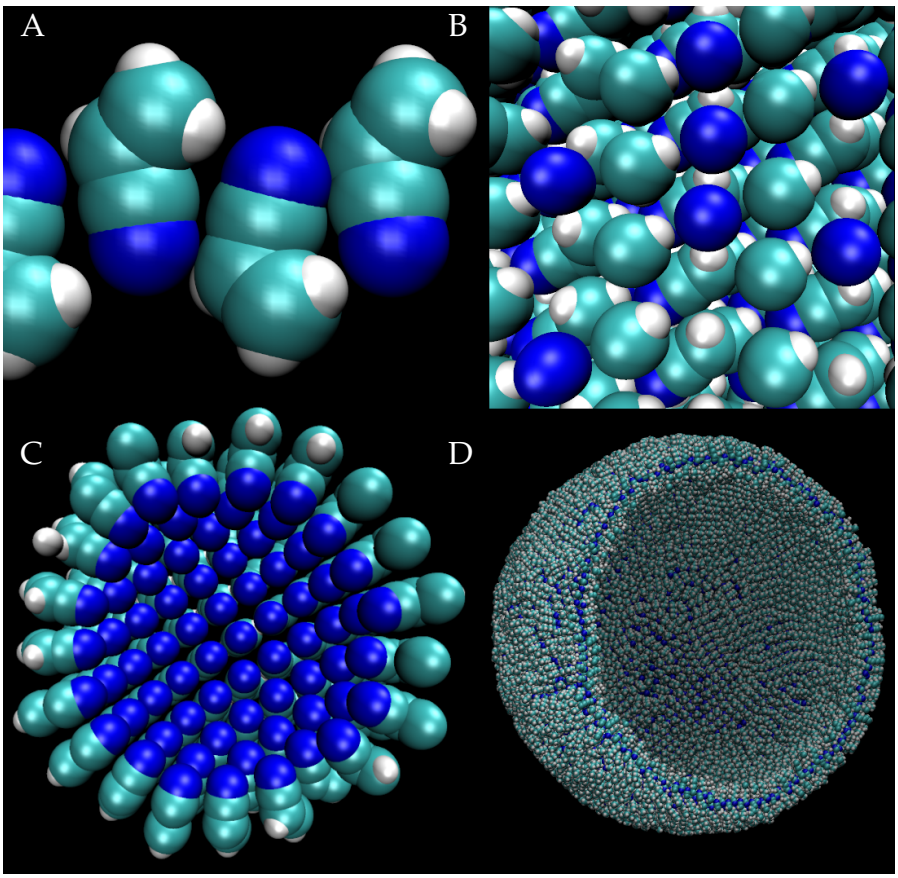


Figure 6.10: States of acrylonitrile. **A)** Azotosome. Interlocking nitrogen and hydrogen atoms reinforce the structure. **B)** Solid. Adjacent nitrogen atoms create some unfavorable repulsion. **C)** Micelle. Adjacent nitrogen atoms make this highly unfavorable. **D)** Azotosome vesicle of diameter 9 nm, the size of a small virus particle.

the real concentrations are currently unknown, so they may or may not be high enough to make azotosomes thermodynamically stable. There is no trend in ΔG with respect to the number of carbon atoms in the chain. ΔG appears to be determined by the way in which the molecules fit together, rather than any simple property of the molecules themselves.

In a cold world without oxygen, we suggest that the vesicles needed for compartmentalization, a key requirement for life, would be very different to those

Species	ΔG (kcal/mol)
Acrylonitrile	7.6
Acetonitrile	17.4
Propanenitrile	6.4
Butanenitrile	7.1
Pentanenitrile	13.4
Hexanenitrile	11.5
Aminopropane	5.9
Aminobutane	6.7
Aminopentane	9.6
Aminohexane	6.2

Table 6.18: Gibbs free energy of decomposition: The net mechanical work required to remove a molecule from the membrane, within 20% uncertainty. These values are concentration-dependent.

found on earth. Rather than long-chain nonpolar molecules that form the prototypical terrestrial membrane in aqueous solution, we find membranes which form in liquid methane at cryogenic temperatures do so from the attraction between polar heads of short-chain molecules that are rich in nitrogen. We have termed such a membrane an azotosome. Surprisingly, we find that the flexibility of such membranes is roughly the same as those of membranes formed in aqueous solutions. Despite the huge difference in temperatures between cryogenic azotosomes and room-temperature terrestrial liposomes, which would make almost any molecular structure rigid, they exhibit surprisingly and excitingly similar responses to mechanical stress.

Based on our criteria of thermodynamic stability, or at least metastability, the azotosome appears to be a realizable cryogenic membrane. Starting from all the known molecular components in the atmosphere of such a world, Titan, we were able to select a couple of candidate molecules which were capable of exhibiting properties which appear to be important for vesicle formation. For example, an acrylonitrile azotosome has good thermodynamic stability, a high

energy barrier to decomposition, and an area expansion modulus similar to that of phospholipid cell membranes in oxygen-rich solutions. Acrylonitrile exists in Titan's atmosphere at a concentration of 10 ppm, and could plausibly be formed on any celestial body with a nitrogen-methane atmosphere.

The availability of molecules with an ability to form cell membranes does not by itself demonstrate that life is possible. But it does direct our search for exotic metabolic and reproductive chemistries that would be similarly compatible under cryogenic conditions. As our understanding of conditions that could nurture extra-terrestrial life expands, so does our probability of finding it, perhaps within a highly unlikely solvent.

The data reported in this chapter, and the custom code used to create it, are located at the repositories https://github.com/jminuse/azotosome-paper and https://github.com/jminuse/titan-solvation-paper.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

The theoretical basis of solution-processed solar cell research is far from mature, but the investigations presented here have made progress toward that goal. In the field of PbS quantum dots, the research presented in the previous chapters has answered the important question of the nature of the surface passivation of PbS quantum dots, which takes the form of separable Pb carboxylate hydrate complexes. In the process of these studies, we discovered previously unknown isomerization states of lead acetate hydrate and predicted the existence of a new *dimer* state in nonpolar solvent, highly interesting for a compound like lead acetate which has been known for thousands of years. With regard to the later phases of the PbS quantum dot reaction, we created the first full all-atom classical force field for PbS quantum dots and showed, using the reactive Molecular Dynamics simulations which our new model allowed, the nature of the reaction barrier in the system, the binding of the surface complexes which must be moved so as to uncover the reacting dot cores. This also explains the reason for the great excess of Pb in the reactants which is necessary for the Hines synthesis to produce monodisperse dots: without ample excess Pb, some dot surfaces will lack full passivation, creating inconsistent reactivity which could lead to differing dot sizes. Control of this process will improve the quality of the dots produced by the Hines synthesis and help experimenters to scale the reaction to larger batches and flow reactors.

For the hybrid perovskites, we have developed the first comprehensive theory of the HOIP solvents, with important applications for perovskite solution processing and eventual manufacturing. We have disproved the existing hy-

potheses in the field which suggested that the solubility in a given solvent, and the strength of the complexes with a given solvent, aligns with the solvent's relative polarity or Hansen solubility parameters. We replaced these incorrect hypotheses with a quick first-principles screening test for solvent complexation based on the physical ability to accept dative bonds. This is easily calculated using the solvent's Mayer bond order. Using larger simulations, we have sampled the conformational space of the Pb complexes and demonstrated the existence of eightfold and ninefold coordination, when the literature suggested only sixfold coordination. This large amount of solvent per complex implies that concentrated perovskite solutions may contain free solvent only as a minor component between networks of interacting complexes. These corrections and new discoveries will help researchers engineer better synthesis processes and better solar cells.

In terms of methylammonium additives for the HOIP synthesis, we have shown that the previously accepted mechanism in the literature, solubility enhancement by a reverse common-ion effect, is most likely not responsible for the solubility enhancement observed for methylammonium ions. We have suggested an alternative mechanism by which methylammonium ions increase the solubility of the lead halides in poor polar solvents, by creating stabilized complexes with the lead halide solutes. We have also discovered that complexes with mixed solvents show nonlinear properties when the two solvents are both high in solubility and linear properties when one of the solvents is poor, due most likely to the dative bonding of the strong solvents affecting the electronic state of the central Pb ion. These discoveries will make it easier to choose new solvents and create sophisticated mixed solvent blends so as to optimize the time/temperature profile and "thermal budget" of the perovskite synthesis. It

will also help to reduce dewetting and thus the formation of pinholes and poor films. Finally, this work has applications beyond the HOIP precursor solutions, because better understanding of the solvent complexation will inform how to control the solid solvent complexes and improve the post-processing techniques which use solvent vapor.

Along with the quantum dot and perovskite solutions, we also examined the solutions which occur in the seas of Saturn's moon Titan. We performed a thorough analysis of the solubility of polar compounds in Titan's nonpolar fluids. We discovered that Titan's naturally-occurring nitrogen compounds dissolve less in its surface fluids than insoluble minerals such as quartz dissolve in water on Earth. This implies that the nitrogen compounds will form hard mineral-like deposits distinct from Titan's primarily hydrocarbon surface. It also implies that Titan's seas, composed largely of methane and nitrogen, will have very low concentrations of any polar compound. Finally, we calculated the properties of possible membranes formed from nitrogen compounds. We found these membranes to have excellent stabilities and elastic properties similar to those of cell membranes on Earth, despite being possible at cryogenic temperatures.

In the process of investigating the solubility effects of methylammonium ions on the perovskite precursors, we illuminated potential structures of the perovskite "monomer" which forms in solution in the presence of methylammonium ions. We have also shown how these perovskite monomers can begin to merge into dimers, passivated by solvent, and so begin the perovskite synthesis reaction mechanism. The study of the later stages of HOIP synthesis, as with the later stages of the PbS quantum dot synthesis, will require a classical force field. The current form of the SMRFF potential is not necessarily suited for this

system because PbX_2 and the perovskite lattice appear to have angle dependent character which the PbS quantum dots do not. Just as the Morse potential is the simplest reactive *pairwise* potential, the simplest reactive potential which allows angle-dependence is the Tersoff potential. A Tersoff functional form is therefore a good starting point for the creation of a highly-in-demand HOIP synthesis force field.

The first order of business for future quantum dot solar cell work must be more Molecular Dynamics and DFT matching of X-ray diffraction data, in order to rationalize the experimental results newly being discovered using the Cornell High-Energy Synchotron Source (CHESS). Much of this data uses alternate compositions from PbS, meaning that in terms of force fields, substitutions such as Cd (for Pb) and Se (for S) must be considered. Faster force field parameterization using a new LAMMPS module we have are developing, "min_style params," makes these compositional changes more viable from the perspective of lengthy force field parameterizations. Quantum dots composed of perovskite crystals are a further compositional change which is of interest, once again requiring a perovskite force field to be studied using MD.

Future perovskite solar cell work based on the studies presented here should first continue the analysis of HOIP synthesis solvents. Ionic liquids, such as the methylammonium formate used to synthesize HOIPs by Professor Lara Estroff (Dept. of Materials Science and Engineering, Cornell University), are an interesting case. The ionic nature of these solvents means that their mechanism of solvation may be quite different from that of the solvents we have studied so far. If this mechanism could be understood at the *ab initio* level, it might illuminate deeper principles of the perovskite solutions.

Methylamine, the solvent used to regenerate and "heal" perovskite layers, must be even more unusual in its solvation mechanism that methylammonium formate, since methylamine contains no obvious binding sites at all: it has no oxygen and its amine group is only a weak Lewis base for coordination bonding due to its two hydrogen atoms. Methylamine has not yet been used as a HOIP synthesis solvent - its quick evaporation and high vapor pressure make it difficult to handle when mixing the reactants, forcing the reaction to be performed at very low temperature or under methylamine gas. However, the use of methylamine as a repair agent for forming smooth perovskite layers implies that its mechanism of solvation is important and its use as main solvent cannot be ruled out. It is also possible that chemical modifications might decrease the vapor pressure without modifying the unexplained but helpful solvation properties of the solvent. For example, ethylamine or propylamine might be compromise replacements for methylamine.

Another subject of consideration is the protic solvents. So far no published HOIP synthesis uses a protic solvent, for reasons which are unclear. What, if anything, makes the protic solvents ineffective? Since humidity has sometimes been found helpful for HOIP synthesis methods, could water be used as a solvent, perhaps in mixtures or with the use of solvation-increasing additives? Or does hydrogen bonding, or the presence of H^+, ruin the HOIP reaction mechanism?

The HOIP reaction mechanism itself must be studied further in order to determine the effect of conditions beyond the initial solutions. We have calculated some HOIP monomer and dimer geometries in our polar organic solvents, but we have not studied the activation energy which links the monomer and dimer

states, nor how it varies with solvent composition and concentration. These factors are vital to a full explanation of the perovskite nucleation.

In the long term, absent breakthroughs in the speed of *ab initio* methods, HOIP synthesis cannot be studied fully without a custom classical force field. Such a force field would allow the study of complex solute-solute effects in the HOIP solutions, especially HOIP colloidal structures, heats of reaction, reaction mechanisms, and activation energies for larger sizes up to the full film.

The revolution in solution-processed solar cell materials has only just begun. As discussed in the introduction, our increasing ability to manufacture solar cells from organic liquid solutions will have far-reaching effects on the energy economy if it reaches the market and further reduces the cost of solar installations. Because we are in the early stages of solution-processed solar cell optimization, trial-and-error methods still predominate over theory, and this situation may be expected to continue for some time to come. However, computational studies like the ones described in this dissertation are building a foundation of theoretical understanding beneath the new experimental discoveries. Our growing knowledge provides the opportunity that laboratory synthesis may be standardized into true manufacturing, making real the immense potential of solution-processed solar cells.

BIBLIOGRAPHY

[1] Vishal Shah, Jerimiah Booream-Phelps, and Susie Min. Outlook: Let the second gold rush begin. *Deutsche Bank: Markets Research*, 2014.

[2] Hawaii State Energy Office. Hawaii energy facts and figures, May 2015. *State of Hawaii Department of Business, Economic Development, and Tourism*, page 6, 2015.

[3] Vishal Shah and Jerimiah Booream-Phelps. Crossing the chasm: Solar grid parity in a low oil price era. *Deutsche Bank: Markets Research*, 2015.

[4] Danylo Zherebetskyy, Marcus Scheele, Yingjie Zhang, Noah Bronstein, Christopher Thompson, David Britt, Miquel Salmeron, Paul Alivisatos, and Lin-Wang Wang. Hydroxylation of the surface of PbS nanocrystals passivated with oleic acid. *Science*, 344(6190):1380–1384, 2014.

[5] Edward H Sargent. Colloidal quantum dot solar cells. *Nature Photonics*, 6 (3):133–135, 2012.

[6] Xiaoqin Li, Yanwen Wu, Duncan Steel, D Gammon, TH Stievater, DS Katzer, D Park, C Piermarocchi, and LJ Sham. An all-optical quantum gate in a semiconductor quantum dot. *Science*, 301(5634):809–811, 2003.

[7] Shuang Jiao, Qing Shen, Iván Mora-Seró, Jin Wang, Zhenxiao Pan, Ke Zhao, Yuki Kuga, Xinhua Zhong, and Juan Bisquert. Band engineering in core/shell znte/cdse for photovoltage and efficiency enhancement in exciplex quantum dot sensitized solar cells. *ACS Nano*, 9(1):908–915, 2015.

[8] Delia J Milliron. Quantum dot solar cells: The surface plays a core role. *Nature Materials*, 13(8):772–773, 2014.

[9] Nicholas C Anderson, Mark P Hendricks, Joshua J Choi, and Jonathan S Owen. Ligand exchange and the stoichiometry of metal chalcogenide nanocrystals: spectroscopic observation of facile metal-carboxylate displacement and binding. *Journal of the American Chemical Society*, 135(49): 18536–18548, 2013.

[10] Matthew C Beard, Octavi E Semonin, Arthur J Nozik, Aaron G Midgett, and Joseph M Luther. Multiple exciton generation in PbSe quantum dots and quantum dot solar cells. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2012.

[11] Sarah Kurtz and Greg Wilson. Best research-cell efficiencies May 2016.

[12] Margaret A Hines and Gregory D Scholes. Colloidal PbS nanocrystals with size-tunable near-infrared emission: observation of post-synthesis self-narrowing of the particle size distribution. *Advanced Materials*, 15(21): 1844–1849, 2003.

[13] Josef Eisinger. Lead and wine. Eberhard Gockel and the colica Pictonum. *Medical History*, 26(03):279–302, 1982.

[14] Philip G. Harrison and Andrew T. Steel. Lead(ii) carboxylate structures. *Journal of Organometallic Chemistry*, 239(1):105 – 113, 1982.

[15] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant. PubChem: Lead diacetate. *Annual Reports in Computational Chemistry*, 4, 2008.

[16] Svetlana V Kilina, Dmitri S Kilin, and Oleg V Prezhdo. Breaking the

phonon bottleneck in pbse and cdse quantum dots: time-domain density functional theory of charge carrier relaxation. *Acs Nano*, 3(1):93–99, 2008.

[17] Patrick R Brown, Donghun Kim, Richard R Lunt, Ni Zhao, Moungi G Bawendi, Jeffrey C Grossman, and Vladimir Bulovic. Energy level modification in lead sulfide quantum dot thin films through ligand exchange. *ACS nano*, 8(6):5863–5872, 2014.

[18] Kaifu Bian, Joshua J Choi, Ananth Kaushik, Paulette Clancy, Detlef-M Smilgies, and Tobias Hanrath. Shape-anisotropy driven symmetry transformations in nanocrystal superlattice polymorphs. *ACS nano*, 5(4):2815–2823, 2011.

[19] Zhaochuan Fan, Li-Chiang Lin, Wim Buijs, Thijs JH Vlugt, and Marijn A van Huis. Atomistic understanding of cation exchange in pbs nanocrystals using simulations with pseudoligands. *Nature communications*, 7, 2016.

[20] Jian Gong, Seth B Darling, and Fengqi You. Perovskite photovoltaics: life-cycle assessment of energy and environmental impacts. *Energy & Environmental Science*, 8(7):1953–1968, 2015.

[21] Ibraheem Almansouri, Anita Ho-Baillie, and Martin A Green. Ultimate efficiency limit of single-junction perovskite and dual-junction perovskite/silicon two-terminal devices. *Japanese Journal of Applied Physics*, 54(8S1):08KD04, 2015.

[22] Yuanyuan Zhou, Onkar S Game, Shuping Pang, and Nitin P Padture. Microstructures of organometal trihalide perovskites for solar cells: their

evolution from solutions and characterization. *Journal of Physical Chemistry Letters*, 6(23):4827–4839, 2015.

[23] Wanyi Nie, Hsinhan Tsai, Reza Asadpour, Jean-Christophe Blancon, Amanda J Neukirch, Gautam Gupta, Jared J Crochet, Manish Chhowalla, Sergei Tretiak, Muhammad A Alam, et al. High-efficiency solution-processed perovskite solar cells with millimeter-scale grains. *Science*, 347 (6221):522–525, 2015.

[24] Giles E Eperon, Victor M Burlakov, Pablo Docampo, Alain Goriely, and Henry J Snaith. Morphological control for high performance, solution-processed planar heterojunction perovskite solar cells. *Advanced Functional Materials*, 24(1):151–157, 2014.

[25] Wei Chen, Yongzhen Wu, Youfeng Yue, Jian Liu, Wenjun Zhang, Xudong Yang, Han Chen, Enbing Bi, Islam Ashraful, Michael Grätzel, et al. Efficient and stable large-area perovskite solar cells with inorganic charge extraction layers. *Science*, 350(6263):944–948, 2015.

[26] Tomas Leijtens, Giles E Eperon, Nakita K Noel, Severin N Habisreutinger, Annamaria Petrozza, and Henry J Snaith. Stability of metal halide perovskite solar cells. *Advanced Energy Materials*, 5(20), 2015.

[27] Xiong Li, M Ibrahim Dar, Chenyi Yi, Jingshan Luo, Manuel Tschumi, Shaik M Zakeeruddin, Mohammad Khaja Nazeeruddin, Hongwei Han, and Michael Grätzel. Improved performance and stability of perovskite solar cells by crystal crosslinking with alkylphosphonic acid ω-ammonium chlorides. *Nature Chemistry*, 2015.

[28] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon JL Billinge,

and Jonathan S Owen. Structure of methylammonium lead iodide within mesoporous titanium dioxide: active material in high-performance perovskite solar cells. *Nano letters*, 14(1):127–133, 2013.

[29] Wan-Jian Yin, Ji-Hui Yang, Joongoo Kang, Yanfa Yan, and Su-Huai Wei. Halide perovskite materials for solar cells: a theoretical review. *Journal of Materials Chemistry A*, 3(17):8926–8942, 2015. Lots of DFT on perovskites. Without dispersion correction, GGA PBE predicts lattice poorly, implying vDW is structural important. DFT predicts perovskite phase transitions poorly. Band gaps also do not follow DFT trends with respect to changing cation. Electron effective mass and hole effective mass calculated, helps explain electrical properties. Looked at disorder, grain boundaries, dangling bonds with DFT: grain boundaries are electrically benign. Studied mixed-halide perovskites, found them to have more stable enthalpy of formation.

[30] Carlo Motta, Fedwa El-Mellouhi, Sabre Kais, Nouar Tabet, Fahhad Alharbi, and Stefano Sanvito. Revealing the role of organic cations in hybrid halide perovskite ch3nh3pbi3. *Nature communications*, 6, 2015. Effect of orientation of MA. Gaussian, periodic BCs, HSE06 with def2. Also, VASP plane waves, PBE with DFT-TS dispersion correction.

[31] Jonathon S Bechtel, Ram Seshadri, and Anton Van der Ven. Energy landscape of molecular motion in cubic methylammonium lead iodide from first principles. *The Journal of Physical Chemistry C*, 2016.

[32] Christopher Eames, Jarvist M Frost, Piers RF Barnes, Brian C Oregan, Aron Walsh, and M Saiful Islam. Ionic transport in hybrid lead iodide perovskite solar cells. *Nature communications*, 6, 2015. Studied activation

energies for ionic migration using DFT to explore hysteresis. Used PBEsol and VASP. 4 4 4 supercell (768 atoms).

[33] Jarvist M Frost, Keith T Butler, Federico Brivio, Christopher H Hendon, Mark Van Schilfgaarde, and Aron Walsh. Atomistic origins of high-performance in hybrid halide perovskite solar cells. *Nano letters*, 14(5): 2584–2590, 2014. Electronic structure calculations: discovered spontaneous electric polarization in HOIPs, studied ferroelectric and antiferroelectric regions to see if different film structure could create higher open circuit voltage. Used PBEsol and VASP.

[34] David T Moore, Kwan W Tan, Hiroaki Sai, Katherine P Barteau, Ulrich Wiesner, and Lara A Estroff. Direct crystallization route to methylammonium lead iodide perovskite from an ionic liquid. *Chemistry of Materials*, 27(9):3197–3199, 2015.

[35] Nam Joong Jeon, Jun Hong Noh, Woon Seok Yang, Young Chan Kim, Seungchan Ryu, Jangwon Seo, and Sang Il Seok. Compositional engineering of perovskite materials for high-performance solar cells. *Nature*, 517 (7535):476–480, 2015.

[36] Seogjoon Yoon, Kevin G Stamplecoskie, and Prashant V Kamat. How lead halide complex chemistry dictates the composition of mixed halide perovskites. *Journal of Physical Chemistry Letters*, pages 1368–1373, 2016.

[37] Namyoung Ahn, Dae-Yong Son, In-Hyuk Jang, Seong Min Kang, Mansoo Choi, and Nam-Gyu Park. Highly reproducible perovskite solar cells with average efficiency of 18.3% and best efficiency of 19.7% fabricated *via* Lewis base adduct of lead (ii) iodide. *Journal of the American Chemical Society*, 137(27):8696–8699, 2015.

[38] Keyou Yan, Mingzhu Long, Tiankai Zhang, Zhanhua Wei, Haining Chen, Shihe Yang, and Jianbin Xu. Hybrid halide perovskite solar cell precursors: Colloidal chemistry and coordination engineering behind device processing for high efficiency. *Journal of the American Chemical Society*, 137 (13):4460–4468, 2015.

[39] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

[40] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. *Chemical Reviews*, 112(1):289–320, 2011.

[41] Kieron Burke. Perspective on density functional theory. *Journal of Chemical Physics*, 136(15):150901, 2012.

[42] Ralf Stowasser and Roald Hoffmann. What do the Kohn-Sham orbitals and eigenvalues mean? *Journal of the American Chemical Society*, 121(14): 3414–3420, 1999.

[43] Shuqiang Niu and Michael B Hall. Theoretical studies on reactions of transition-metal complexes. *Chemical Reviews*, 100(2):353–406, 2000.

[44] Lars Goerigk and Stefan Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13(14):6670–6688, 2011.

[45] Christopher J Cramer and Donald G Truhlar. Density functional theory for transition metals and transition metal chemistry. *Physical Chemistry Chemical Physics*, 11(46):10757–10816, 2009.

[46] Yihan Shao, Laszlo Fusti Molnar, Yousung Jung, Jörg Kussmann, Christian Ochsenfeld, Shawn T Brown, Andrew TB Gilbert, Lyudmila V Slipchenko, Sergey V Levchenko, Darragh P ONeill, et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics*, 8(27):3172–3191, 2006.

[47] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

[48] Frank H Stillinger and Thomas A Weber. Computer simulation of local order in condensed phases of silicon. *Physical Review B*, 31(8):5262, 1985.

[49] Jerry Tersoff. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B*, 39(8):5566, 1989.

[50] Byeong-Joo Lee and MI Baskes. Second nearest-neighbor modified embedded-atom-method potential. *Physical Review B*, 62(13):8564, 2000.

[51] Murray S Daw, Stephen M Foiles, and Michael I Baskes. The embedded-atom method: a review of theory and applications. *Materials Science Reports*, 9(7):251–310, 1993.

[52] Donald W Brenner. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Physical Review B*, 42(15):9458, 1990.

[53] Adri Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. ReaxFF: a reactive force field for hydrocarbons. *Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[54] Tao Liang, Yun Kyung Shin, Yu-Ting Cheng, Dundar E Yilmaz, Karthik Guda Vishnu, Osvalds Verners, Chenyu Zou, Simon R Phillpot,

Susan B Sinnott, and Adri CT van Duin. Reactive potentials for advanced atomistic simulations. *Annual Review of Materials Research*, 43:109–129, 2013.

[55] Y Xiao, W Dong, and HF Busnengo. Reactive force fields for surface chemical reactions: A case study with hydrogen dissociation on Pd surfaces. *Journal of Chemical Physics*, 132(1):014704, 2010.

[56] DG Pettifor and II Oleinik. Analytic bond-order potentials beyond Tersoff-Brenner. i. theory. *Physical Review B*, 59(13):8487, 1999.

[57] DK Ward, XW Zhou, Bryan Matthew Wong, F Patrick Doty, and Jonathan A Zimmerman. Analytical bond-order potential for the Cd-Zn-Te ternary system. *Physical Review B*, 86(24):245203, 2012.

[58] Clive R Bealing, William J Baumgardner, Joshua J Choi, Tobias Hanrath, and Richard G Hennig. Predicting nanocrystal shape through consideration of surface-ligand interactions. *ACS Nano*, 6(3):2118–2127, 2012.

[59] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.

[60] John P Perdew, Kieron Burke, and Yue Wang. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Physical Review B*, 54(23):16533, 1996.

[61] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. Quantum espresso: a modular and

open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009.

[62] Frank Neese. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.

[63] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, and GA Petersson. Gaussian 09, Rev. B. 01. *Gaussian Inc., Wallingford, CT*, 2010.

[64] David E Woon and Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon. *Journal of Chemical Physics*, 98(2):1358–1371, 1993.

[65] Artur F Izmaylov, Gustavo E Scuseria, and Michael J Frisch. Efficient evaluation of short-range Hartree-Fock exchange in large molecules and periodic systems. *Journal of Chemical Physics*, 125(10):104103, 2006.

[66] Vincent L Chevrier, Shyue Ping Ong, Rickard Armiento, Maria KY Chan, and Gerbrand Ceder. Hybrid density functional calculations of redox potentials and formation energies of transition metal compounds. *Physical Review B*, 82(7):075122, 2010.

[67] Thomas M Henderson, Artur F Izmaylov, Giovanni Scalmani, and Gustavo E Scuseria. Can short-range hybrids describe long-range-dependent properties? *Journal of Chemical Physics*, 131(4):044108, 2009.

[68] Walter J Stevens, Morris Krauss, Harold Basch, and Paul G Jasien. Relativistic compact effective potentials and efficient, shared-exponent basis sets for the third-, fourth-, and fifth-row atoms. *Canadian Journal of Chemistry*, 70(2):612–630, 1992.

[69] P Jeffrey Hay and Willard R Wadt. *Ab initio* effective core potentials for molecular calculations. potentials for K to Au including the outermost core orbitals. *Journal of Chemical Physics*, 82(1):299–310, 1985.

[70] Andreas Bergner, Michael Dolg, Wolfgang Küchle, Hermann Stoll, and Heinzwerner Preuß. *Ab initio* energy-adjusted pseudopotentials for elements of groups 13–17. *Molecular Physics*, 80(6):1431–1441, 1993.

[71] Hongwei Gao, Xiujuan Wei, Xuting Liu, and Tingxia Yan. Comparison of different theory models and basis sets in the calculations of structures and 13c NMR spectra of [Pt (en)(CBDCA- O, O)], an analogue of the antitumor drug carboplatin. *Journal of Physical Chemistry B*, 114(11):4056–4062, 2010.

[72] José-Zeferino Ramírez, Rubicelia Vargas, Jorge Garza, and Benjamin P Hay. Performance of the effective core potentials of Ca, Hg, and Pb in complexes with ligands containing N and O donor atoms. *Journal of Chemical Theory and Computation*, 2(6):1510–1519, 2006.

[73] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3094, 2005.

[74] Florian Weigend. Accurate coulomb-fitting basis sets for H to Rn. *Physical Chemistry Chemical Physics*, 8(9):1057–1065, 2006.

[75] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B*, 113(18):6378–6396, 2009.

[76] Feliu Maseras and Keiji Morokuma. IMOMM: A new integrated *ab initio*+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry*, 16(9): 1170–1179, 1995.

[77] Hannes Jónsson, Greg Mills, and Karsten W Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. 1998.

[78] Chunyang Peng, Philippe Y Ayala, H Bernhard Schlegel, and Michael J Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry*, (17): 49–56, 1996.

[79] WF Van Gunsteren and HJC Berendsen. Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *Journal of computer-aided molecular design*, 1(2):171–176, 1987.

[80] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of chemical physics*, 128(14):144120, 2008.

[81] Pedro Salvador, Béla Paizs, Miquel Duran, and Sándor Suhai. On the effect of the bsse on intermolecular potential energy surfaces. comparison of a priori and a posteriori bsse correction schemes. *Journal of Computational Chemistry*, 22(7):765–786, 2001.

[82] Holger Kruse and Stefan Grimme. A geometrical correction for the inter- and intra-molecular basis set superposition error in hartree-fock and den-

sity functional theory calculations for large systems. *Journal of Chemical Physics*, 136(15):154101, 2012.

[83] Friedrich Becker and Alfred Maelicke. Thermokinetische messungen nach dem prinzip der wärmeflusskalorimetrie. *Zeitschrift für Physikalische Chemie*, 55(5_6):280–295, 1967.

[84] Louis A Girifalco and Victor G Weizer. Application of the Morse potential function to cubic metals. *Physical Review*, 114(3):687, 1959.

[85] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins *via* comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[86] Yan Zhao and Donald G Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Physics*, 125(19): 194101, 2006.

[87] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics*, 110(13):6158–6170, 1999.

[88] Thomas M Henderson, Artur F Izmaylov, Gustavo E Scuseria, and Andreas Savin. Assessment of a middle-range hybrid functional. *Journal of Chemical Theory and Computation*, 4(8):1254–1262, 2008.

[89] Martin Head-Gordon and Teresa Head-Gordon. Analytic MP2 frequencies without fifth-order storage. theory and application to bifurcated hy-

drogen bonds in the water hexamer. *Chemical Physics Letters*, 220(1):122–128, 1994.

[90] Gustavo E Scuseria, Curtis L Janssen, and Henry F Schaefer Iii. An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *Journal of Chemical Physics*, 89(12):7382–7387, 1988.

[91] John G Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3(5):300–313, 1935.

[92] Rafal Bogacz and Jonathan D Cohen. Parameterization of connectionist models. *Behavior Research Methods, Instruments, & Computers*, 36(4):732–741, 2004.

[93] Steven G Johnson. NLopt nonlinear-optimization package. Joannopoulos Research Group, Massachusetts Institute of Technology: Cambridge, MA, 2015.

[94] Giacomo Fiorin, Jérôme Hénin, and Axel Kohlmeyer. Collective variables module reference manual for LAMMPS. Sandia National Laboratories: Albuquerque, NM, 2013.

[95] Steve Plimpton, Paul Crozier, and Aidan Thompson. LAMMPS large-scale atomic/molecular massively parallel simulator, version 15 may 2015. Sandia National Laboratories: Albuquerque, NM, 2015.

[96] Nam Joong Jeon, Jun Hong Noh, Young Chan Kim, Woon Seok Yang, Seungchan Ryu, and Sang Il Seok. Solvent engineering for high-performance inorganic–organic hybrid perovskite solar cells. *Nature materials*, 13(9):897–903, 2014.

[97] Zhengguo Xiao, Cheng Bi, Yuchuan Shao, Qingfeng Dong, Qi Wang, Yongbo Yuan, Chenggong Wang, Yongli Gao, and Jinsong Huang. Efficient, high yield perovskite photovoltaic devices grown by interdiffusion of solution-processed precursor stacking layers. *Energy & Environmental Science*, 7(8):2619–2623, 2014.

[98] Chang Liu, Kai Wang, Chao Yi, Xiaojun Shi, Adam W Smith, Xiong Gong, and Alan J Heeger. Efficient perovskite hybrid photovoltaics *via* alcohol-vapor annealing treatment. *Advanced Functional Materials*, 26(1):101–110, 2016.

[99] Peter J Holliman, Eurig W Jones, Arthur Connell, Sanjay Ghosh, Leo Furnell, and Robert J Hobbs. Solvent issues during processing and device lifetime for perovskite solar cells. *Materials Research Innovations*, 19(7):508–511, 2015.

[100] B Selin Tosun and Hugh W Hillhouse. Enhanced carrier lifetimes of pure iodide hybrid perovskite via vapor-equilibrated re-growth (verg). *The journal of physical chemistry letters*, 6(13):2503–2508, 2015.

[101] Ting Zhao, Spencer T Williams, Chu-Chen Chueh, Po-Wei Liang, David S Ginger, Alex K-Y Jen, et al. Design rules for the broad application of fast (¡ 1 s) methylamine vapor based, hybrid perovskite post deposition treatments. *RSC Advances*, 6(33):27475–27484, 2016.

[102] Bin Yang, Ondrej Dyck, Jonathan Poplawsky, Jong Keum, Sanjib Das, Alexander Puretzky, Tolga Aytug, Pooran C Joshi, Christopher M Rouleau, Gerd Duscher, et al. Controllable growth of perovskite films by room-temperature air exposure for efficient planar heterojunction photo-

voltaic cells. *Angewandte Chemie International Edition*, 54(49):14862–14865, 2015.

[103] Xiu Gong, Meng Li, Xiao-Bo Shi, Heng Ma, Zhao-Kui Wang, and Liang-Sheng Liao. Controllable perovskite crystallization by water additive for high-performance solar cells. *Advanced Functional Materials*, 25(42):6671–6678, 2015.

[104] Christian Reichardt. *Solvents and Solvent Effects in Organic Chemistry, 3rd Ed.*, page 418. Wiley-VCH, 2003.

[105] Adam J Bridgeman, Germán Cavigliasso, Luke R Ireland, and Joanne Rothery. The Mayer bond order as a tool in inorganic chemistry. *Journal of the Chemical Society, Dalton Transactions*, (14):2095–2108, 2001.

[106] Adam J Bridgeman and Germán Cavigliasso. Towards an understanding of the bonding in polyoxometalates through bond order and bond energy analysis. *Faraday Discussions*, 124:239–258, 2003.

[107] Robert S Mulliken. Bonding power of electrons and theory of valence. *Chemical Reviews*, 9(3):347–388, 1931.

[108] Ruven L Davidovich, Vitalie Stavila, Dmitry V Marinin, Elena I Voit, and Kenton H Whitmire. Stereochemistry of lead (ii) complexes with oxygen donor ligands. *Coordination Chemistry Reviews*, 253(9):1316–1352, 2009.

[109] Yongping Fu, Fei Meng, Matthew B Rowley, Blaise J Thompson, Melinda J Shearer, Dewei Ma, Robert J Hamers, John C Wright, and Song Jin. Solution growth of single crystal methylammonium lead halide perovskite nanostructures for optoelectronic and photovoltaic applications. *Journal of the American Chemical Society*, 137(17):5810–5818, 2015.

[110] Oscar E. Lanford and Samuel J. Kiehl. The solubility of lead iodide in solutions of potassium iodide-complex lead iodide ions. *Journal of the American Chemical Society*, 63(3):667–669, 1941. doi: 10.1021/ja01848a010.

[111] Letian Dou, Andrew B Wong, Yi Yu, Minliang Lai, Nikolay Kornienko, Samuel W Eaton, Anthony Fu, Connor G Bischak, Jie Ma, Tina Ding, et al. Atomically thin two-dimensional organic-inorganic hybrid perovskites. *Science*, 349(6255):1518–1521, 2015.

[112] Wei Zhang, Michael Saliba, David T Moore, Sandeep K Pathak, Maximilian T Hörantner, Thomas Stergiopoulos, Samuel D Stranks, Giles E Eperon, Jack A Alexander-Webber, Antonio Abate, et al. Ultrasmooth organic–inorganic perovskite thin-film formation and crystallization for efficient planar heterojunction solar cells. *Nature Communications*, 6:6142, 2015.

[113] Lars Goerigk and Stefan Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13(14):6670–6688, 2011.

[114] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics*, 132(15):154104, 2010.

[115] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13): 2157–2164, 2009.

[116] Charles M Hansen. *Hansen solubility parameters: a user's handbook*. CRC press, 2007.

[117] Steven Abbott and Charles M Hansen. *Hansen solubility parameters in practice*, page 36. Hansen-Solubility, 2008.

[118] Derek YC Chan, D John Mitchell, and Barry W Ninham. A model of solvent structure around ions. *The Journal of Chemical Physics*, 70(6):2946–2957, 1979.

[119] Daniel Cordier, Olivier Mousis, Jonathan I Lunine, Panayotis Lavvas, and Véronique Vuitton. An estimate of the chemical composition of Titan's lakes. *The Astrophysical Journal Letters*, 707(2):L128, 2009.

[120] Marco Mastrogiuseppe, Valerio Poggiali, Alexander Hayes, Ralph Lorenz, Jonathan Lunine, Giovanni Picardi, Roberto Seu, Enrico Flamini, Giuseppe Mitri, Claudia Notarnicola, Philippe Paillou, and Howard Zebker. The bathymetry of a Titan sea. *Geophysical Research Letters*, 41(5): 1432–1437, 2014.

[121] MH Langhans, R Jaumann, K Stephan, RH Brown, BJ Buratti, RN Clark, KH Baines, PD Nicholson, RD Lorenz, LA Soderblom, et al. Titan's fluvial valleys: Morphology, distribution, and spectral properties. *Planetary and Space Science*, 60(1):34–51, 2012.

[122] Chao He and Mark A Smith. Solubility and stability investigation of Titan aerosol analogs: New insight from NMR analysis. *Icarus*, 232:54–59, 2014.

[123] Ingo Müller-Wodarg, Caitlin A Griffith, Emmanuel Lellouch, and Thomas E Cravens. *Titan: Interior, Surface, Atmosphere, and Space Environment*, volume 14. Cambridge University Press, 2014.

[124] Christopher A Strulson, Rosalynn C Molden, Christine D Keating, and Philip C Bevilacqua. RNA catalysis through compartmentalization. *Nature Chemistry*, 4(11):941–946, 2012.

[125] J. P. Ferris and D. A. Usher. Origins of Life. In G. Zubay, editor, *Biochemistry*, chapter 32. Addison-Wesley, 1983.

[126] Christopher McKay and Heather Smith. Possibilities for methanogenic life in liquid methane on the surface of Titan. *Icarus*, 178:274–276, 2005.

[127] D. Strobel. Molecular hydrogen in Titan's atmosphere: Implications of the measured tropospheric and thermospheric mole fractions. *Icarus*, 208: 878–886, 2010.

[128] Roger N. Clark, John M. Curchin, Jason W. Barnes, Ralf Jaumann, Larry Soderblom, Dale P. Cruikshank, Robert H. Brown, Sbastien Rodriguez, Jonathan Lunine, Katrin Stephan, Todd M. Hoefen, Stphane Le Moulic, Christophe Sotin, Kevin H. Baines, Bonnie J. Buratti, and Philip D. Nicholson. Detection and mapping of hydrocarbon deposits on Titan. *Journal of Geophysical Research: Planets*, 115(E10), 2010.

[129] HB Niemann, SK Atreya, JE Demick, D Gautier, JA Haberman, DN Harpold, WT Kasprzak, JI Lunine, TC Owen, and F Raulin. Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini-Huygens probe gas chromatograph mass spectrometer experiment. *Journal of Geophysical Research: Planets (1991–2012)*, 115(E12), 2010.

[130] S. Singh, V. F. Chevrier, M. Leitner, M. Gainor, and L. Roe. Solubility of organics in liquid hydrocarbons under Titan surface conditions. *45th Lunar and Planetary Science Conference*, 2014.

[131] A. Neumann and R. Mann. Die lslichkeit von festem acetylen in flssigen methan/thylen-mischungen. *Chemie Ingenieur Technik*, 41(12):708–711, 1969. ISSN 1522-2640.

[132] Sugata P Tan, Hertanto Adidharma, Jeffrey S Kargel, and Giles M Marion. Equation of state for solid solution–liquid–vapor equilibria at cryogenic conditions. *Fluid Phase Equilibria*, 360:320–331, 2013.

[133] Feliciano M. Llave, Kraemer D. Luks, and James P. Kohn. Three-phase liquid-liquid-vapor equilibria in the nitrogen-methane-ethane and nitrogen-methane-propane systems. *Journal of Chemical and Engineering Data*, 32(1):14–17, 1987.

[134] G Trappehl and H Knapp. Vapour-liquid equilibria in the ternary mixtures $N_2CH_4C_2H_6$ and $N_2C_2H_6C_3H_8$. *Cryogenics*, 27(12):696 – 716, 1987.

[135] David Harry Gabis. Liquidvapor equilibrium at 90160 K for systems containing nitrogen, methane, and ethane. Master's thesis, Cornell University, Ithaca, NY, 1991.

[136] Christopher R Glein and Everett L Shock. A geochemical model of non-ideal solutions in the methane–ethane–propane–nitrogen–acetylene system on Titan. *Geochimica et Cosmochimica Acta*, 115:217–240, 2013.

[137] Michael J Malaska and Robert Hodyss. Dissolution of benzene, naphthalene, and biphenyl in a simulated Titan lake. *Icarus*, 242:74–81, 2014.

[138] G. T. Preston and J. M. Prausnitz. Thermodynamics of solid solubility in cryogenic solvents. *Industrial and Engineering Chemistry Process Design and Development*, 9(2):264–271, 1970.

[139] Daniel Cordier, Jason W. Barnes, and Abel Ferreira. Composition of Titan's dry lakebeds: what can be inferred from the solubility theory. *44th Lunar and Planetary Science Conference*, 2013.

[140] A Luspay-Kuti, VF Chevrier, D Cordier, EG Rivera-Valentin, S Singh, A Wagner, and FC Wasiak. Experimental constraints on the composition and dynamics of Titan's polar lakes. *Earth and Planetary Science Letters*, 410:75–83, 2015.

[141] George Firanescu, David Luckhaus, Grenfell N. Patey, Sushil K. Atreya, and Ruth Signorell. The composition of liquid methanenitrogen aerosols in Titans lower atmosphere from Monte Carlo simulations. *Icarus*, 212(2): 779 – 789, 2011.

[142] Ding-Yu Peng and Donald B Robinson. A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals*, 15(1):59–64, 1976.

[143] Giorgio Soave. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chemical Engineering Science*, 27(6):1197–1203, 1972.

[144] Andreas Klamt and GCOSMO Schüürmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, (5):799–805, 1993.

[145] Andreas Klamt and Frank Eckert. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilibria*, 172(1):43–72, 2000.

[146] Frank Eckert and Andreas Klamt. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal*, 48(2):369–385, 2002.

[147] G Te Velde, F Matthias Bickelhaupt, Evert Jan Baerends, C Fonseca Guerra, Stan JA van Gisbergen, Jaap G Snijders, and Tom Ziegler. Chemistry with ADF. *Journal of Computational Chemistry*, 22(9):931–967, 2001.

[148] C Fonseca Guerra, JG Snijders, G Te Velde, and EJ Baerends. Towards an order-N DFT method. *Theoretical Chemistry Accounts*, 99(6):391–403, 1998.

[149] N. Sarker, A. Somogyi, J. I. Lunine, and M. A. Smith. Titan aerosol analogues: analysis of the nonvolatile tholins. *Astrobiology*, 3(4):719–726, 2003.

[150] Walter G Chapman, George Jackson, and Keith E Gubbins. Phase equilibria of associating fluids: chain molecules with multiple bonding sites. *Molecular Physics*, 65(5):1057–1079, 1988.

[151] Walter G Chapman, Keith E Gubbins, George Jackson, and Maciej Radosz. New reference equation of state for associating liquids. *Industrial & Engineering Chemistry Research*, 29(8):1709–1721, 1990.

[152] MS Wertheim. Fluids with highly directional attractive forces. i. statistical thermodynamics. *Journal of Statistical Physics*, 35(1-2):19–34, 1984.

[153] MS Wertheim. Fluids with highly directional attractive forces. ii. thermodynamic perturbation theory and integral equations. *Journal of Statistical Physics*, 35(1-2):35–47, 1984.

[154] MS Wertheim. Fluids with highly directional attractive forces. iii. multiple attraction sites. *Journal of Statistical Physics*, 42(3-4):459–476, 1986.

[155] MS Wertheim. Fluids with highly directional attractive forces. iv. equilibrium polymerization. *Journal of Statistical Physics*, 42(3-4):477–492, 1986.

[156] MS Wertheim. Thermodynamic perturbation theory of polymerization. *The Journal of Chemical Physics*, 87(12):7323–7331, 1987.

[157] Sharon G Sauer and Walter G Chapman. A parametric study of dipolar chain theory with applications to ketone mixtures. *Industrial & Engineering Chemistry Research*, 42(22):5687–5696, 2003.

[158] Prasanna K Jog and WG Chapman. Application of Wertheim's thermodynamic perturbation theory to dipolar hard sphere chains. *Molecular Physics*, 97(3):307–319, 1999.

[159] Prasanna K Jog, Sharon G Sauer, Jorg Blaesing, and Walter G Chapman. Application of dipolar chain theory to the phase behavior of polar fluids and mixtures. *Industrial & Engineering Chemistry Research*, 40(21):4641–4648, 2001.

[160] EJS Gomes de Azevedo and JCG Calado. Thermodynamics of liquid methane + ethane. *Fluid Phase Equilibria*, 49:21–34, 1989.

[161] Jorge CG Calado, Gerald A Garcia, and Lionel AK Staveley. Thermodynamics of the liquid system methane + propane. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 70: 1445–1451, 1974.

[162] Xiao Hong Han, Yu Jia Zhang, Zan Jun Gao, Ying Jie Xu, Qin Wang, and Guang Ming Chen. Vapor–liquid equilibrium for the mixture nitrogen (N_2)+ methane (CH_4) in the temperature range of (110 to 125) K. *Journal of Chemical & Engineering Data*, 57(5):1621–1626, 2012.

[163] Robert Hodyss, Mathieu Choukroun, Christophe Sotin, and Patricia Beauchamp. The solubility of 40 Ar and 84 Kr in liquid hydrocarbons:

Implications for Titan's geological evolution. *Geophysical Research Letters*, 40(12):2935–2940, 2013.

[164] Elzbieta Szczepaniec-Cieciak. Solubility of solids and liquids in cryogenic liquids. *The Experimental Determination of Solubilities, Volume 6*, pages 557–593, 2003.

[165] André B de Haan, Rainer Bölts, and Jürgen Gmehling. Vapor-liquid equilibria and excess enthalpies for binary mixtures of acrylonitrile with hexane, cyclohexane, benzene, toluene, 2-butanone, and acetonitrile. *Journal of Chemical & Engineering Data*, 41(5):1155–1159, 1996.

[166] Jean Pierre Monfort. Vapor-liquid equilibriums for benzene-acetonitrile and toluene-acetonitrile mixtures at 343.15 K. *Journal of Chemical and Engineering Data*, 28(1):24–27, 1983.

[167] Maria Antosik, Maria Galka, and Stanislaw K Malanowski. Vapor-liquid equilibrium for acetonitrile+ propanenitrile and 1-pentanamine + 1-methoxy-2-propanol. *Journal of Chemical & Engineering Data*, 49(1):11–17, 2004.

[168] Svetlana Miroshnichenko, Thomas Grützner, Daniel Staak, and Jadran Vrabec. Molecular simulation of the vapor–liquid phase behavior of cyanides and their binary mixtures. *Fluid Phase Equilibria*, 354:286–297, 2013.

[169] Johan Aaqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry*, 94(21):8021–8024, 1990.

[170] Scott A. Best, Kenneth M. Merz, and Charles H. Reynolds. Free energy perturbation study of octanol/water partition coefficients: comparison with continuum GB/SA calculations. *The Journal of Physical Chemistry B*, 103(4):714–726, 1999.

[171] Stephen E. DeBolt and Peter A. Kollman. Investigation of structure, dynamics, and solvation in 1-octanol and its water-saturated solution: Molecular dynamics and free-energy perturbation studies. *Journal of the American Chemical Society*, 117(19):5316–5340, 1995.

[172] TP Straatsma and HJC Berendsen. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of Chemical Physics*, 89(9):5876–5886, 1988.

[173] Volkhard Helms and Rebecca C Wade. Free energies of hydration from thermodynamic integration: comparison of molecular mechanics force fields and evaluation of calculation accuracy. *Journal of Computational Chemistry*, 18(4):449–462, 1997.

[174] Nuno M. Garrido, Antnio J. Queimada, Miguel Jorge, Eugnia A. Macedo, and Ioannis G. Economou. 1-octanol/water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies. *Journal of Chemical Theory and Computation*, 5(9):2436–2446, 2009.

[175] George Kaminski, Erin M. Duffy, Tooru Matsui, and William L. Jorgensen. Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *The Journal of Physical Chemistry*, 98(49): 13077–13082, 1994.

[176] W. Jorgensen. *OPLS All-Atom Parameters for Organic Molecules, Ions, Peptides and Nucleic Acids*. Yale University, New Haven, CT, 2009.

[177] J.P. Bouanich. Site-site Lennard-Jones potential parameters for N_2, O_2, H_2, CO and CO_2. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 47 (4):243 – 250, 1992.

[178] Pavel Hobza, Martin Kabelac, Jiri Sponer, Petr Mejzlik, and Jiri Vondrasek. Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree-Fock results. *Journal of Computational Chemistry*, 18(9):1136–1150, 1997.

[179] Tomáš Boublík. Hard-sphere equation of state. *The Journal of Chemical Physics*, 53(1):471–472, 1970.

[180] GA Mansoori, NF Carnahan, KE Starling, and TW Leland Jr. Equilibrium thermodynamic properties of the mixture of hard spheres. *The Journal of Chemical Physics*, 54(4):1523–1525, 1971.

[181] John A Barker and Douglas Henderson. Perturbation theory and equation of state for fluids. ii. a successful theory of liquids. *The Journal of Chemical Physics*, 47(11):4714–4721, 1967.

[182] Joachim Gross and Gabriele Sadowski. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Industrial & Engineering Chemistry Research*, 40(4):1244–1260, 2001.

[183] Joachim Gross and Gabriele Sadowski. Application of perturbation theory

to a hard-chain reference fluid: an equation of state for square-well chains. *Fluid Phase Equilibria*, 168(2):183–199, 2000.

[184] Jeffrey P Wolbach and Stanley I Sandler. Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures. *Industrial & Engineering Chemistry Eesearch*, 37(8):2917–2928, 1998.

[185] Otto Redlich and JNS Kwong. On the thermodynamics of solutions. V. an equation of state. fugacities of gaseous solutions. *Chemical Reviews*, 44(1): 233–244, 1949.

[186] Daniel Cordier, Olivier Mousis, Jonathan I Lunine, Panayotis Lavvas, and Véronique Vuitton. An estimate of the chemical composition of Titan's lakes. *The Astrophysical Journal Letters*, 707(2):L128, 2009.

[187] Sugata P Tan, Jeffrey S Kargel, and Giles M Marion. Titans atmosphere and surface liquid: New calculation using Statistical Associating Fluid Theory. *Icarus*, 222(1):53–72, 2013.

[188] Christopher R Glein and Everett L Shock. A geochemical model of non-ideal solutions in the methane–ethane–propane–nitrogen–acetylene system on Titan. *Geochimica et Cosmochimica Acta*, 115:217–240, 2013.

[189] RL Rowley, WV Wilding, JL Oscarson, Y Yang, NA Zundel, TE Daubert, and RP Danner. DIPPR data compilation of pure chemical properties. *Design Institute for Physical Properties*, 2006.

[190] Carl L Yaws. *Handbook of Chemical Compound Data for Process Safety*. Gulf Professional Publishing, 1997.

[191] DR Lide. *2003–2004 CRC Handbook of Chemistry and Physics*. CRC Press, 2003.

[192] Rolf Prydz and Robert D Goodwin. Experimental melting and vapor pressures of methane. *The Journal of Chemical Thermodynamics*, 4(1):127–133, 1972.

[193] J Dojcansky and J Heinrich. Saturated vapour pressure of acetonitrile. *Chem. Zvesti*, 28(2):157–159, 1974.

[194] N.N. Sevrugova, V.A. Sokorskii, and N.M. Zhavoronkov. Phase equilibrium for mixtures acrylonitrile-acetonitrile. *Zh. Prikl. Khim*, 37(9):1989–1993, 1964.

[195] Ch B Willingham, WJ Taylor, Joan M Pignocco, and FD Rossini. Vapor pressures and boiling points of some paraffin, alkylcyclopentane, alkylcyclohexane, and alkylbenzene hydrocarbons. *Journal of Research of the National Bureau of Standards*, 35(3):219–244, 1945.

[196] Grant F Carruth and Riki Kobayashi. Vapor pressure of normal paraffins ethane through n-decane from their triple points to about 10 mm mercury. *Journal of Chemical and Engineering Data*, 18(2):115–126, 1973.

[197] Heima Sinozaki, Ryozaburo Hara, and Shinroku Mitsukuri. The vapour pressures of hydrogen cyanide. *Bulletin of the Chemical Society of Japan*, 1 (4):59–61, 1926.

[198] Merardo P Edejer and George Thodos. Vapor pressures of liquid nitrogen between the triple and critical points. *Journal of Chemical and Engineering Data*, 12(2):206–209, 1967.

[199] JD Kemp and Clark J Egan. Hindered rotation of the methyl groups in propane. the heat capacity, vapor pressure, heats of fusion and vaporiza-

tion of propane. entropy and density of the gas. *Journal of the American Chemical Society*, 60(7):1521–1525, 1938.

[200] RR Dreisbach and SA Shrader. Vapor pressure–temperature data on some organic compounds. *Industrial & Engineering Chemistry*, 41(12):2879–2880, 1949.

[201] Georgios M Kontogeorgis and Georgios K Folas. *Thermodynamic models for industrial applications: from classical and advanced mixing rules to association theories*. John Wiley & Sons, 2009.

[202] Aleksandra Dominik, Walter G Chapman, Matthias Kleiner, and Gabriele Sadowski. Modeling of polar systems with the perturbed-chain SAFT equation of state. investigation of the performance of two polar terms. *Industrial & Engineering Chemistry Research*, 44(17):6928–6938, 2005.

[203] Nandou Lu, David A Kofke, and Thomas B Woolf. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *Journal of Computational Chemistry*, 25(1):28–40, 2004.

[204] Martin Walker, Andrew J. A. Harvey, Ananya Sen, and Caroline E. H. Dessent. Performance of M06, M06-2X, and M06-HF density functionals for conformationally flexible anionic clusters: M06 functionals perform better than B3LYP for a model system with dispersion and ionic hydrogen-bonding interactions. *The Journal of Physical Chemistry A*, 117 (47):12590–12600, 2013.

[205] J. A. Plumley and J. J. Dannenberg. A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer

with emphasis on basis set superposition error. *J Comput. Chem.*, 32(8): 1519–1527, 2011.

[206] Xiaosong Li and Michael J. Frisch. Energy-represented direct inversion in the iterative subspace within a hybrid geometry optimization method. *Journal of Chemical Theory and Computation*, 2(3):835–839, 2006.

[207] J. Tomasi, B. Mennucci, and R. Cammi. Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3093, 2005.

[208] Mark E Tuckerman, Jos Alejandre, Roberto Lpez-Rendn, Andrea L Jochim, and Glenn J Martyna. A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermalisobaric ensemble. *Journal of Physics A: Mathematical and General*, 39 (19):5629, 2006.

[209] M Fulchignoni, F Ferri, F Angrilli, AJ Ball, A Bar-Nun, MA Barucci, C Bettanini, G Bianchini, W Borucki, G Colombatti, et al. In situ measurements of the physical characteristics of Titan's environment. *Nature*, 438(7069): 785–791, 2005.

[210] E.L. Pollock and Jim Glosli. Comments on P3M, FMM, and the Ewald method for large periodic Coulombic systems. *Computer Physics Communications*, 95(23):93 – 110, 1996.

[211] Matt Yarrison, Kenneth R Cox, and Walter G Chapman. Measurement and modeling of the solubility of water in supercritical methane and ethane from 310 to 477 K and pressures from 3.4 to 110 MPa. *Industrial & Engineering Chemistry Research*, 45(20):6770–6777, 2006.

[212] H Langer and H Offermann. On the solubility of sodium chloride in water. *Journal of Crystal Growth*, 60(2):389–392, 1982.

[213] GW Morey, RO Fournier, and JJ Rowe. The solubility of quartz in water in the temperature interval from 25 to 300 C. *Geochimica et Cosmochimica Acta*, 26(10):1029–1043, 1962.

[214] Leslie E Orgel. The origin of lifea review of facts and speculations. *Trends Biochem. Sci.*, 23(12):491–495, 1998.

[215] Christopher A Strulson, Rosalynn C Molden, Christine D Keating, and Philip C Bevilacqua. Rna catalysis through compartmentalization. *Nat. Chem.*, 4(11):941–946, 2012.

[216] Daniel Segré, Dafna Ben-Eli, David W Deamer, and Doron Lancet. The lipid world. *Orig. Life Evol. Biosph.*, 31(1-2):119–145, 2001.

[217] David Deamer. A giant step towards artificial life? *Trends Biotechnol.*, 23 (7):336–338, 2005.

[218] Juli Peretó, Purificación López-García, and David Moreira. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.*, 29(9): 469–477, 2004.

[219] Peter Walde. Surfactant assemblies and their various possible roles for the origin (s) of life. *Orig. Life Evol. Biosph.*, 36(2):109–150, 2006.

[220] Hironobu Kunieda, Kazuki Shigeta, and Masao Suzuki. Phase behavior and formation of reverse vesicles in long-polyoxyethylene-chain nonionic surfactant systems. *Langmuir*, 15(9):3118–3122, 1999.

[221] Henk Mollee, Jeroen De Vrind, and Tom De Vringer. Stable reversed vesicles in oil: Characterization studies and encapsulation of model compounds. *J. Pharm. Sci.*, 89(7):930–939, 2000.

[222] Hongguang Li, Jingcheng Hao, and Zhonghua Wu. Phase behavior and properties of reverse vesicles in salt-free catanionic surfactant mixtures. *J. Phys. Chem. B*, 112(12):3705–3710, 2008.

[223] Shih-Huang Tung, Hee-Young Lee, and Srinivasa R Raghavan. A facile route for creating reverse vesicles: insights into reverse self-assembly in organic liquids. *J. Am. Chem. Soc.*, 130(27):8813–8817, 2008.

[224] Lucy H Norman. Is there life on titan? *Astron. Geophys.*, 52(1):1–39, 2011.

[225] Ingo Müller-Wodarg, Caitlin A Griffith, Emmanuel Lellouch, and Thomas E Cravens. *Titan: Interior, Surface, Atmosphere, and Space Environment*, volume 14. Cambridge University Press, 2014.

[226] CP McKay and HD Smith. Possibilities for methanogenic life in liquid methane on the surface of titan. *Icarus*, 178(1):274–276, 2005.

[227] Darrell F Strobel. Molecular hydrogen in titans atmosphere: Implications of the measured tropospheric and thermospheric mole fractions. *Icarus*, 208(2):878–886, 2010.

[228] Roger N Clark, John M Curchin, Jason W Barnes, Ralf Jaumann, Larry Soderblom, Dale P Cruikshank, Robert H Brown, Sébastien Rodriguez, Jonathan Lunine, Katrin Stephan, et al. Detection and mapping of hydrocarbon deposits on titan. *J. Geophys. Res.: Planets*, 115(E10), 2010.

[229] Dirk Schulze-Makuch, Alberto G Fairén, and Alfonso Davila. Locally

targeted ecosynthesis: A proactive in situ search for extant life on other worlds. *Astrobiology*, 13(7):674–678, 2013.

[230] Niladri Sarker, Arpad Somogyi, Jonathan I Lunine, and Mark A Smith. Titan aerosol analogues: analysis of the nonvolatile tholins. *Astrobiology*, 3(4):719–726, 2003.

[231] Christofer Hofsäß, Erik Lindahl, and Olle Edholm. Molecular dynamics simulations of phospholipid bilayers with cholesterol. *Biophys. J.*, 84(4): 2192–2206, 2003.

[232] Shamik Sen, Shyamsundar Subramanian, and Dennis E Discher. Indentation and adhesive probing of a cell membrane with afm: theoretical model and experiments. *Biophys. J.*, 89(5):3203–3213, 2005.

[233] Gary Ayton, Alexander M Smondyrev, Scott G Bardenhagen, Patrick McMurtry, and Gregory A Voth. Calculating the bulk modulus for a lipid bilayer with nonequilibrium molecular dynamics simulation. *Biophys. J.*, 82(3):1226–1238, 2002.

[234] DD Lasic. On the thermodynamic stability of liposomes. *J. Colloid Interface Sci.*, 140(1):302–304, 1990.

[235] P Van der Ploeg and HJC Berendsen. Molecular dynamics simulation of a bilayer membrane. *J. Chem. Phys.*, 76(6):3271–3276, 1982.

[236] Richard M Venable, Yuhong Zhang, Barry J Hardy, and Richard W Pastor. Molecular dynamics simulations of a lipid bilayer and of hexadecane: an investigation of membrane fluidity. *Science*, 262(5131):223–226, 1993.

[237] Rüdiger Goetz, Gerhard Gompper, and Reinhard Lipowsky. Mobility and elasticity of self-assembled membranes. *Phys. Rev. Lett.*, 82(1):221, 1999.

[238] Dennis E Discher, Vanessa Ortiz, Goundla Srinivas, Michael L Klein, Younghoon Kim, David Christian, Shenshen Cai, Peter Photos, and Fariyal Ahmed. Emerging applications of polymersomes in delivery: from molecular dynamics to shrinkage of tumors. *Prog. Polym. Sci.*, 32 (8):838–857, 2007.

[239] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45): 11225–11236, 1996.

[240] Yandi Hu and Young-Shin Jun. Biotite dissolution in brine at varied temperatures and co2 pressures: its activation energy and potential co2 intercalation. *Langmuir*, 28(41):14633–14641, 2012.

[241] W. Jorgensen. *OPLS All-Atom Parameters for Organic Molecules, Ions, Peptides & Nucleic Acids*. Yale University, 2009.

[242] Mark E Tuckerman, José Alejandre, Roberto López-Rendón, Andrea L Jochim, and Glenn J Martyna. A liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble. *J. Phys. A: Math. Gen.*, 39(19):5629, 2006.

[243] EL Pollock and Jim Glosli. Comments on p¡ sup¿ 3¡/sup¿ m, fmm, and the ewald method for large periodic coulombic systems. *Comput. Phys. Commun.*, 95(2):93–110, 1996.

[244] Yoshihiro Yokoyama and Yuji Ohashi. Crystal structure of acrylonitrile. *Bull. Chem. Soc. Jpn.*, 71(2):345–348, 1998.

[245] Daniel Sheppard and Graeme Henkelman. Paths to which the nudged elastic band converges. *Journal of Computational Chemistry*, 32(8):1769–1771, 2011.

[246] Daniel Sheppard, Rye Terrell, and Graeme Henkelman. Optimization methods for finding minimum energy paths. *Journal of Chemical Physics*, 128(13):134106, 2008.

[247] Daniel Sheppard, Penghao Xiao, William Chemelewski, Duane D Johnson, and Graeme Henkelman. A generalized solid-state nudged elastic band method. *Journal of Chemical Physics*, 136(7):074103, 2012.

[248] Dominic R Alfonso and Kenneth D Jordan. A flexible nudged elastic band program for optimization of minimum energy pathways using *ab initio* electronic structure methods. *Journal of Computational Chemistry*, 24(8): 990–996, 2003.

[249] David W Eggert, Adele Lorusso, and Robert B Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9(5-6):272–290, 1997.

[250] Vebjørn Bakken and Trygve Helgaker. The efficient optimization of molecular geometries using redundant internal coordinates. *Journal of Chemical Physics*, 117(20):9160–9174, 2002.

[251] Marcus D Hanwell, Donald Ephraim Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, 4(1):17, 2012.

APPENDIX A

A NEW ALGORITHM FOR NON-PERIODIC NUDGED ELASTIC BAND

The Nudged Elastic Band (NEB) algorithm is one of the most robust tools for the analysis of chemical transition states. However, in non-periodic systems, it has an undesirable feature: for every image in the path, there are six excess degrees of freedom which have no physical meaning, representing the non-existent "translation and rotation" of each image. We propose an elegant means of removing these degrees of freedom within the NEB algorithm using the orthogonal Procrustes method.

In the limit of many images, NEB is guaranteed to converge to the Steepest Descent Path (related to, but not necessarily the same as, the minimum energy path) given a suitable optimization algorithm. [245] The choice of optimization algorithms is outside the scope of this work, but many good options exist, with a global Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm being preferred in the literature. [246]

As of this writing, there is limited competition for the Nudged Elastic Band family of algorithms. Alternatives exist, such as the string method, but they tend to converge more slowly than NEB in most cases and thus are not widely used. [246] Instead, the field has focused on new entries to the NEB family. Doubly-Nudged Elastic Band (DNEB) modifies NEB by adding some spring force perpendicular to the path, with the effect of shortening the path. [246] This modification speeds up the early steps, but tends to impede long-run convergence, especially for curved paths - DNEB has been mathematically proved to not converge properly is the optimum path is curved [246]. The Climbing-Image NEB algorithm (CI-NEB) adds a good follow-up to an ordinary NEB calculation

by allowing one image near the saddle point to increase in energy, thereby reducing the number of images necessary to resolve the highest point. Of course, this requires the saddle point to be found with reasonable accuracy first. [246] In short, NEB and its close relatives dominate the field.

However, there is an interesting and hitherto unresolved contradiction within this success. It has long been observed that, in NEB, the position and orientation of each image with respect to the others has no physical significance. The only chemically relevant variables are the internal coordinates within each image: in the Cartesian representation, six excess degrees of freedom exist. In periodic NEB, these excess degrees of freedom can be removed by constraining the periodic lattice. [247] In non-periodic NEB, this is not an option, and other methods have been employed to remove these excess degrees of freedom, without practical success. For example, one atom may be constrained to a point, a second to a line, and a third to a plane. [248] This certainly removes the excess degrees of freedom, but the selection of the atoms is arbitrary and the constraint interferes sharply with the original forces on the system. [248]

Our insight is that *the space in which the images interact is arbitrary*. The spring forces between adjacent images can be calculated in *any* Cartesian basis without interfering with the assumptions behind NEB. From this insight, it is only a small step to the idea that these bases should be selected deliberately to promote the convergence of the algorithm. Much of NEB is based around the idea of spring forces competing with the real physical forces on the system. If possible, it would be desirable to minimize the spring forces, so as to allow the physical forces their greatest effect. This we proceed to do.

A.1 Methods

The orthogonal Procrustes algorithm [249] minimizes the Frobenius norm between two sets of points by means of translation and rigid rotation. The Frobenius norm is equivalent to the sum of the Euclidean distances, which are directly proportional to the spring forces, so the use of the orthogonal Procrustes algorithm is equivalent to minimizing the spring forces. Therefore, the orthogonal Procrustes algorithm uniquely determines a basis for interaction of the NEB images. Furthermore, this basis appears to be superior to any other basis in that it minimizes the non-physical spring forces.

The Procrustes rotation can be applied to the NEB path in this way:

$$x'_i = P(x_i, x'_{i-1}) \text{ for } i = 2, 3..., n \tag{A.1}$$

In which x_i is image i out of n images on the path, x'_i is the rotated image which will be used to determine the spring forces, and $P(a, b)$ indicates the Procrustes rotation of image a to match image b. Starting from image 1 implies that this is the image to which all others will be rotated; this choice is arbitrary because it does not impact the final spring distances, apart from the small differences due to limited floating point precision.

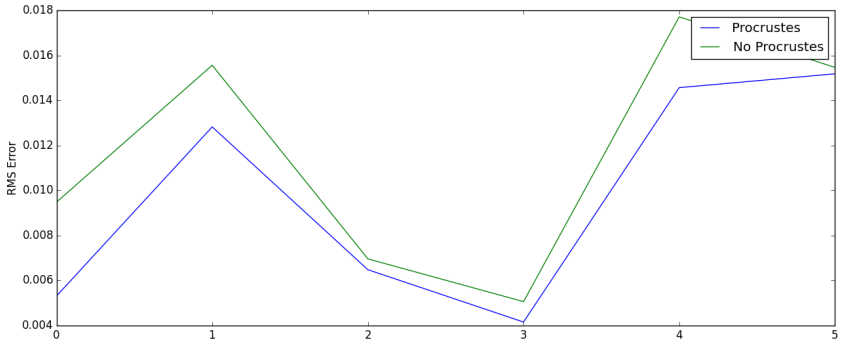
In the same spirit, we remove the net translation forces from the gradient. Since the Procrustes algorithm will remove all translation, these forces no longer have meaning, forming ”noise” on top of the ”signal” of the structure-deforming forces. There is no unique way to remove the net forces; our method is to sum the total net force on each image’s atom and then divide it equally

among all the atoms and subtract it out.

A.2 Results

Unfortunately, for reasons that are not yet clear, this Procrustes modification of the NEB algorithm does not produce any notable improvement. The negative result is shown in Figure A.1. It is possible that this is due to implementation errors (NEB and Procrustes both being complex algorithms). Alternatively, it may be that the convergence of the underlying optimization algorithm is hindered by the implicit rotations, which might cause the curvature information stored by the algorithm to be incorrect.

Figure A.1: NEB error with and without Procrustes rotation (see inset). Unfortunately, Procrustes rotation seems to make the error worse, not better.



Bolstering this idea, the removal of net translation forces taken by itself is effective at improving the NEB performance. It is therefore possible that the removal of net rotational forces (net torque) could also be beneficial. Removal of torque is more mathematically complex, but still possible. The forces must be transformed into a basis of internal coordinates, in which no net force or torque

can exist, and then transformed back out into Cartesian coordinates. This has the effect of purging the net force and torque. The matrix which applies the change of basis between internal coordinates q_i and Cartesian coordinates x_j is the Wilson B matrix [250]:

$$B_{ij} = \frac{\delta q_i}{\delta x_j} \tag{A.2}$$

Generally "redundant" internal coordinates q are used, in which the dimension of q is greater than that of x, making the Wilson B matrix rectangular.

As an alternative to optimizing in Cartesian coordinates and purging them of net force and torque at each step, the optimization could be performed in internal coordinates and the spring forces could be converted into internal coordinates, also using the Wilson B matrix. Molecular geometry optimization in internal coordinates generally performs better than the same optimization in Cartesian coordinates, probably because it is easier to make an accurate initial guess at the Hessian using internal coordinates [250]. NEB in internal coordinates might therefore be a valuable improvement to the algorithm. Surprisingly, this idea has never been implemented, perhaps because it requires the additional step of Procrustes rotation in order to create a basis for the spring forces.

Since the NEB family of algorithms leads the field of transition path finding, it is well worth considering ideas to make NEB optimizations more effective. A more efficient NEB variation for non-periodic systems would be of great benefit to computational chemistry.

APPENDIX B

CLANCELOT, A SOFTWARE PACKAGE FOR COMPUTATIONAL CHEMISTRY

The Clancelot package is designed to simplify computational chemistry research by separating domain knowledge about software packages such as Orca from more important and general chemical knowledge. Using Avogadro [251], Packmol [115], LAMMPS [95], and Orca [62] through Clancelot should not require any knowledge of the xyz or cml file formats, the LAMMPS data file format, or Packmol or Orca input or output files. A knowledge of LAMMPS data files is still necessary for a good reason: LAMMPS has enough options (thousands of commands and dozens of packages) that there is no way to encapsulate it without substantial leakage of the abstraction. A leaky abstraction can be worse than none at all because it requires knowledge of *both* the abstraction and the underlying system.

In the future, as Clancelot support is extended to more software packages, this principle should be followed. For example, the periodic DFT package Quantum Espresso [61] uses input files which contain a great deal of information, midway between Orca and LAMMPS in complexity. Unless a means can be found to package Quantum Espresso inputs into an abstraction without leaks, it is better not to attempt to simplify the interface to fit that of Orca or Gaussian.

Useful future work for the Clancelot package includes:

- Quantum Espresso support
- LAMMPS input file support based on arrays of strings (commands), so as not to require explicitly opening and closing a file

- Building in sensible defaults for Orca, *e.g.* "Opt LooseOpt B97-D3 def2-SVP GCP(DFT/SV)" which implies a quick but sensible geometry optimization
- Improve the pre-commit bug tests, making them more portable and faster.
- Consider Continuous Integration, such as Travis CI, which integrates with Github. This is a more sophisticated version of the pre-commit bug tests but may be impossible if the remote server must run LAMMPS, Orca, etc to perform the tests.
- Write new tests before new features (test-driven development)

The code for Clancelot is available online:

https://github.com/jminuse/clancelot.